

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2005, 23:15:55 ; Search time 166 Seconds
(without alignments)
368.121 Million cell updates/sec

Title: US-09-525-041-2

Sequence: 1 MASSRMALLLSLCAKTGV.....LTWSNECKRQHFICKTRP 158

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq16Dec04:*

1: geneseqp1980s:*\n2: geneseqp1990s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	878	100.0	158	AAW12691	Aaw12691 Human col
2	878	100.0	158	AAW37929	Aaw37929 A human R
3	878	100.0	158	AAW37866	Aaw37866 Human pro
4	878	100.0	158	AAW84274	AAW84274 Protein e
5	878	100.0	158	AAW12900	AAW12900 Human col
6	878	100.0	158	AAW74934	AAW74934 Human TSA
7	878	100.0	158	AAW24519	AAW24519 C880P sim
8	878	100.0	158	AAW24520	AAW24520 C880P sim
9	878	100.0	158	AAW24521	AAW24521 C880P sim
10	878	100.0	158	AAW24517	AAW24517 C81-152 C
11	878	100.0	158	AAW24518	AAW24518 C880P sim
12	878	100.0	158	AAW29829	AAW29829 Human REG
13	878	100.0	158	AAW78993	AAW78993 Human REG
14	878	100.0	158	AAW85851	AAW85851 Human can
15	878	100.0	158	AAW60022	AAW60022 Human REG
16	878	100.0	158	AAW83832	AAW83832 Human REG
17	878	100.0	158	AAW55366	AAW55366 Human col
18	878	100.0	158	AAW55370	AAW55370 Human col
19	878	100.0	158	AAW55369	AAW55369 Human col
20	878	100.0	158	AAW55367	AAW55367 Human col
21	878	100.0	158	AAW55368	AAW55368 Human col
22	878	100.0	158	AAW80529	AAW80529 Ovarian c
23	878	100.0	158	AAW75523	AAW75523 Prostate
24	878	100.0	158	AAW16659	AAW16659 Human alb
25	878	100.0	158	AAW56454	AAW56454 Human col

26	878	100.0	158	ADP85474	ADP85474 Human reg
27	878	100.0	158	ADH21872	ADH21872 Human Reg
28	878	100.0	158	ADN39463	ADN39463 Cancer/an
29	878	100.0	158	ADN39882	ADN39882 Cancer/an
30	878	100.0	158	ADN38820	ADN38820 Cancer/an
31	878	100.0	158	ADN39543	ADN39543 Cancer/an
32	878	100.0	158	AD158211	AD158211 Human reg
33	878	100.0	158	ADN35904	ADN35904 Human reg
34	878	100.0	158	ADQ29679	ADQ29679 Human col
35	878	100.0	158	ADQ80359	ADQ80359 Regenerat
36	878	100.0	158	ADP55956	ADP55956 Human PRO
37	878	100.0	166	AAW75620	AAW75620 Human col
38	878	100.0	767	ADP16496	ADP16496 Human alb
39	878	100.0	767	ADH21795	ADH21795 Human alb
40	867	98.7	158	ABP99310	ABP99310 Amino aci
41	856	97.5	153	AAW92267	AAW92267 Human can
42	760	86.6	367	ADP73150	ADP73150 RELP-Fc F
43	695	79.2	122	ABP76307	ABP76307 Human GEN
44	307.5	35.0	134	ABP59097	ABP59097 Human zin
45	254.5	29.0	165	AAW81513	AAW81513 Sequence

ALIGNMENTS

RESULT 1
AAW12691 standard; protein; 158 AA.
AAW12691;
31-MAY-1997 (first entry)
Human colon specific protein.
Human colon specific protein.
Colon specific protein; colon cancer; metastasis; diagnosis; therapy;
antibody; vaccine; agonist; antagonist.
Homo sapiens.
WO9639541-A1.
12-DEC-1996.
06-JUN-1995; 95WO-US007169.
06-JUN-1995; 95WO-US007169.
06-JUN-1995; 95WO-US007169.
(HUMA-) HUMAN GENOME SCI INC.
Soppet DR, Li Y, Dillon PJ;
WPI-4997-043162/04.
N-PSDB; AAT51784.
New isolated colon specific gene - used to develop prods. for use in the
diagnosis and treatment of colon disorders, partic. colon cancer.
Claim 1; Page 53; 64pp; English.
A human colon specific protein (AAW12691) is a potential diagnostic
marker for colon cancer. It is believed that the presence of active
transcription of the colon specific gene in non-colon cells of a host is
indicative of colon cancer metastases. The amino acid sequence of the
colon specific protein was deduced from a cDNA clone (AA151784) isolated
from a human colon cancer cDNA library. Recombinant colon specific
protein can be produced in transformed host (e.g. bacterial, insect)
cells and used to develop prods. for the diagnosis and treatment of colon
disorders, partic. colon cancer metastasis. Antibodies raised against the
protein can be used to target cancer cells and as part of a colon cancer
vaccine
Sequence 158 AA;

Query Match 100.0%; Score 878; DB 2; Length 158;
Best Local Similarity 100.0%; Pred. No. 6.8e-83;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRMRLLLLLSCIAKTGVLDIIMRPSCAPGWFYHKSNCYGFYRKLRMNSDALEECQS 60
DB 1 MASRMRLLLLLSCIAKTGVLDIIMRPSCAPGWFYHKSNCYGFYRKLRMNSDALEECQS 60
QY 61 YGNNGAHLASIIISLKEASTIAEYISGYORSOPITWIGLHPQKROQOWIDGAMLYRSWSG 120
DB 61 YGNNGAHLASIIISLKEASTIAEYISGYORSOPITWIGLHPQKROQOWIDGAMLYRSWSG 120
QY 121 KSMGNKHCALMSNNNFLTWSNRCNRQHFCLKYRP 158
DB 121 KSMGNKHCALMSNNNFLTWSNRCNRQHFCLKYRP 158

RESULT 2
AAM37929

ID AAM37929 standard; protein; 158 AA.

AC AAM37929;

DT 21-AUG-1998 (first entry)

DE A human Reg I-gamma protein.

KM Human; Reg I-gamma protein; C-type lectin; Incyte clone 1310334;

KM regulation; cell growth; development; tumorigenesis; neurodegeneration;

KM inhibition; treatment; prevention; neoplasia; metastasis;

KM neurodegenerative change; Alzheimer's disease; Down's syndrome;

KM regeneration; pancreatic beta-cells; diabetes.

OS Homo sapiens.

XX MO9816640-A1.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US018174.

XX 11-OCT-1996; 96US-00729103.

XX (INCY-) INCYTE PHARM INC.

XX Bandman O, Goli SK;

XX WPI, 1998-251287/22.

XX N-PSDB; AAV29156.

XX New isolated human Reg I-gamma protein useful for developing products

XX for treating, e.g. diabetes, tumours, or neuro-degenerative disease such

XX as Alzheimer's.

XX Claim 1; Fig 1; 72pp; English.

XX The present sequence represents a human Reg I-gamma protein, which

XX comprises a C-type lectin. The sequence was identified in Incyte clone

XX 1310334. Reg I-gamma protein is involved in regulation of cell growth and

XX development. Since the overexpression of reg protein is associated with

XX tumorigenesis and neurodegeneration, inhibition of human Reg I-gamma

XX expression can be used for treating or preventing neoplasia or metastasis

XX and neurodegenerative changes associated with Alzheimer's disease and

XX other disorders of the central nervous system, e.g. Down's syndrome. Reg

XX I-gamma can also be used in therapeutics to induce regeneration of

XX pancreatic beta-cells in the treatment of diabetes. The products can also

XX be used for detection for, e.g. expression of REG I-gamma, diagnosis and

XX drug screen

XX Sequence 158 AA;

XX Query Match 100.0%; Score 878; DB 2; Length 158;

Best Local Similarity 100.0%; Pred. No. 6.8e-83;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRMRLLLLLSCIAKTGVLDIIMRPSCAPGWFYHKSNCYGFYRKLRMNSDALEECQS 60
DB 1 MASRMRLLLLLSCIAKTGVLDIIMRPSCAPGWFYHKSNCYGFYRKLRMNSDALEECQS 60
QY 61 YGNNGAHLASIIISLKEASTIAEYISGYORSOPITWIGLHPQKROQOWIDGAMLYRSWSG 120
DB 61 YGNNGAHLASIIISLKEASTIAEYISGYORSOPITWIGLHPQKROQOWIDGAMLYRSWSG 120
QY 121 KSMGNKHCALMSNNNFLTWSNRCNRQHFCLKYRP 158
DB 121 KSMGNKHCALMSNNNFLTWSNRCNRQHFCLKYRP 158

RESULT 3
AAM37866

ID AAM37866 standard; protein; 158 AA.

AC AAM37866;

DT 10-AUG-1998 (first entry)

DE Human protein comprising secretory signal amino acid sequence 3.

KM Human protein; secretory signal; nutritional source; cytokine; immunity;

KM haematopoiesis; activin; inhibin; tumour; chemotactic; chemokine;

KM thrombolytic; anti-inflammatory; inhibition; stomach cancer cell.

OS Homo sapiens.

XX WO9811217-A2.

XX 19-MAR-1998.

XX 12-SEP-1997; 97WO-JP003239.

XX 13-SEP-1996; 96JP-00243060.

XX (SAGA) SAGAMI CHEM RES CENTRE.

XX (PROT-) PROTEGENE INC.

XX Kato S, Sekine S, Yamaguchi T, Kobayashi M;

XX WPI, 1998-207380/18.

XX N-PSDB; AAV29035; AAV29036.

XX Human proteins with secretory signal sequences - used to treat immune

XX deficiencies, infections, tumours, and haematopoietic disorders, etc.

XX Claim 1; Page 67-68; 13pp; English.

XX This is the amino acid sequence of a novel human protein comprising a

XX secretory signal isolated from stomach cancer cells. Its proteins can be

XX used as nutritional sources or supplements. The proteins may also have

XX cytokine functions, immune modulating functions, haematopoiesis

XX regulating activity, activin/inhibin regulating activity,

XX chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,

XX receptor/ligand activity, anti-inflammatory activity, tumour inhibition

XX activity

XX Sequence 158 AA;

XX Query Match 100.0%; Score 878; DB 2; Length 158;

XX Best Local Similarity 100.0%; Pred. No. 6.8e-83;

XX Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRMRLLLLLSCIAKTGVLDIIMRPSCAPGWFYHKSNCYGFYRKLRMNSDALEECQS 60

DB 1 MASRMRLLLLLSCIAKTGVLDIIMRPSCAPGWFYHKSNCYGFYRKLRMNSDALEECQS 60

QY 61 YGNNGAHLASIIISLKEASTIAEYISGYORSOPITWIGLHPQKROQOWIDGAMLYRSWSG 120

Db 61 YNGGAHLASITLSEKASTIAEYISGYORSQPIWIGLHPQKQOMWIDGAMVLYRSWSG 120
QY 121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLCKYRP 158
Db 121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLCKYRP 158

RESULT 4

AAW84274
ID AAW84274 standard; protein; 158 AA.

AC AAW84274;

DT 25-MAR-1999 (first entry)

DE Protein encoded by a human colon specific gene.

KM Human; colon specific gene; diagnosis; colon disorder; colon cancer;
viability; colon cancer cell.

OS Homo sapiens.

PN US5861494-A.

PD 19-JAN-1999.

PF 06-JUN-1995; 95US-00468413.

PR 06-JUN-1995; 95US-00468413.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Dillon PJ, Soppet DR, Li Y;

DR WPI; 1999-130432/11.

DR N-PSDB; AAX03195.

PT Isolated human colon specific gene - used to develop products for the
diagnosis and treatment of disorders of the colon, e.g. colon cancer and
metastases.

PS Claim 1; Fig 1A-C; 20pp; English.

CC The present sequence is encoded by a human colon specific gene. The
nucleic acid sequence can be used to develop products for the diagnosis
of a disorder of the colon, e.g. colon cancer or metastases. The products
can also be used to screen for agonists or antagonists for the
polypeptides. The antagonists may be used to treat colon cancer, since
they interact with the function of colon specific polypeptides to inhibit
functions which are necessary for the viability of colon cancer cells.
CC The products can also be used for the production of antibodies and for
the identification of receptors for the polypeptides

XX Sequence 158 AA;

Query Match 100.0%; Score 878; DB 2; Length 158;

Best Local Similarity 100.0%; Pred. No. 6.8e-83;

Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRSMRLLILSLCAKTGVLGDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEECOS 60

Db 1 MASRSMRLLILSLCAKTGVLGDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEECOS 60

QY 61 YNGGAHLASITLSEKASTIAEYISGYORSQPIWIGLHPQKQOMWIDGAMVLYRSWSG 120

Db 61 YNGGAHLASITLSEKASTIAEYISGYORSQPIWIGLHPQKQOMWIDGAMVLYRSWSG 120

QY 121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLCKYRP 158

Db 121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLCKYRP 158

RESULT 5

AAAB12900
ID AAAB12900 standard; protein; 158 AA.

AC AAAB12900;

DT 14-NOV-2000 (first entry)

DE Human colon specific protein sequence.

KM Human; colon specific; colon cancer; metastasis; diagnosis; treatment;
cytostatic.

OS Homo sapiens.

PN US6080722-A.

PD 27-JUN-2000.

PF 29-SEP-1998; 98US-00162508.

PR 06-JUN-1995; 95US-00468413.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Dillon PJ, Soppet DR, Li Y;

DR WPI; 2000-464055/40.

DR N-PSDB; AAA62951.

PT Novel human colon specific polypeptides and polynucleotides for diagnosis
and treatment of colon cancer, for screening compounds which interact
with polypeptide, for synthesis of DNA and manufacture of DNA vectors.

PS Claim 1; Fig 1; 20pp; English.

CC This invention relates to a purified human protein, which is primarily
expressed in tissue derived from the colon. The protein is 152 amino
acids in length and exhibits cytostatic activity. The present sequence
represents the amino acid sequence of the colon specific protein. The
CC protein can be used in the diagnosis and treatment of colon cancer, and
it is thought that abnormally high levels of the gene expression in non-
colon cells is an indication of colon cancer metastasis

XX Sequence 158 AA;

Query Match 100.0%; Score 878; DB 3; Length 158;

Best Local Similarity 100.0%; Pred. No. 6.8e-83;

Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRSMRLLILSLCAKTGVLGDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEECOS 60

Db 1 MASRSMRLLILSLCAKTGVLGDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEECOS 60

QY 61 YNGGAHLASITLSEKASTIAEYISGYORSQPIWIGLHPQKQOMWIDGAMVLYRSWSG 120

Db 61 YNGGAHLASITLSEKASTIAEYISGYORSQPIWIGLHPQKQOMWIDGAMVLYRSWSG 120

QY 121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLCKYRP 158

Db 121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLCKYRP 158

RESULT 6

AAAB74934
ID AAAB74934 standard; protein; 158 AA.

AC AAAB74934;

DT 27-JUN-2001 (first entry)

DE Human TSA7005 protein SEQ ID NO:1.

KM Human; TSA7005; Reg; pancreatic beta cell growth; hypoglycaemic;
 XX diagnosis.
 OS Homo sapiens.
 FM JP2001025389-A.
 XX
 XX 30-JAN-2001.
 XX
 XX 15-JUL-1999; 99JP-00201279.
 XX
 XX 15-JUL-1999; 99JP-00201279.
 XX
 XX (SAKA) OTSUKA PHARM CO LTD.
 XX
 DR WPI: 2001-303742/32.
 DR N-ESDB; AAF82117, AAF82118.
 XX
 XX TSA7005 gene, encoding a polypeptide useful for the diagnosis and
 PT treatment of diseases associated with its expression.
 XX
 XX Claim 1; Page 23; 25pp; Japanese.
 PS
 XX The present sequence represents a human TSA7005 protein which shares 32%
 CC homology with human and mouse Reg proteins, and 34% homology with the rat
 CC Reg protein. TSA7005 has pancreatic beta cell growth activity and
 CC hypoglycaemic activity. The TSA7005 protein can be used for the diagnosis
 CC and treatment of diseases associated with the gene and its expression
 CC product
 XX
 SQ Sequence 158 AA;
 Query Match 100.0%; Score 878; DB 4; Length 158;
 Best Local Similarity 100.0%; Pred. No. 6.8e-83;
 Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASRSMRLLLLSCLAKTGVLDIMRPSCAPGWFYHKSNCYGFRLKLRMSDALEECQS 60
 DB 1 MASRSMRLLLLSCLAKTGVLDIMRPSCAPGWFYHKSNCYGFRLKLRMSDALEECQS 60
 QY 61 YGNGAHLASIIISLKEASTIAEYISGYORSQPIWIGLHPQKRQOMWIDGAMLYRSWSG 120
 DB 61 YGNGAHLASIIISLKEASTIAEYISGYORSQPIWIGLHPQKRQOMWIDGAMLYRSWSG 120
 QY 121 KSMGNGKHCAMSSNNNFLTWSNCKRQHFLLCKYRP 158
 DB 121 KSMGNGKHCAMSSNNNFLTWSNCKRQHFLLCKYRP 158
 RESULT 7
 AAM24519
 ID AAM24519 standard; protein; 158 AA.
 AC AAM24519;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE C880P similar amino acid sequence (GENESSEQ W37866).
 XX
 KW Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
 XX gene therapy; vaccine; colonic cancer.
 OS Homo sapiens.
 XX
 XX WO200149716-A2.
 XX
 XX 12-JUL-2001.
 XX
 XX 29-DEC-2000; 2000WO-US0355596.
 XX
 XX 30-DEC-1999; 99US-00476296.
 XX
 XX 10-JAN-2000; 2000US-00480321.
 XX
 XX 15-FEB-2000; 2000US-00504629.

PR 06-MAR-2000; 2000US-00519444.
 PR 19-MAY-2000; 2000US-00575251.
 PR 29-JUN-2000; 2000US-00609448.
 PR 28-AUG-2000; 2000US-00649811.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
 PI King GR, Wang T, Jiang Y;
 XX
 DR WPI: 2001-441847/47.
 XX
 PT Colon tumor associated proteins and nucleic acids useful for the
 XX prevention, diagnosis and treatment of colonic cancer.
 PS
 XX Claim 2; Page 467-468; 472pp; English.
 CC
 CC The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be used
 CC to treat disorders associated with decreased expression by rectifying
 CC mutations or deletions in a patient's genome that affect the activity of
 CC TCAPs by expressing inactive proteins or to supplement the patient's own
 CC production of them. Additionally, (II) may be used to produce the TCAP
 CC proteins, by inserting the nucleic acids into a host cell culturing the
 CC cell to express the protein. (II) and its complementary sequences may
 CC also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
 CC and hybridisation assays to detect and quantitate the presence of similar
 CC nucleic acids in samples, and therefore which patients may be in need of
 CC restorative therapy. (I) may also be used as antigens in the production
 CC of antibodies against TCAPs and in assays to identify modulators of TCAP
 CC expression and activity. Anti-(I) antibodies and antagonists may also be
 CC used to down regulate TCAP expression and activity. The anti-(I)
 CC antibodies may also be used as diagnostic agents for detecting the
 CC presence of TCAPs in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA)). AA128460 to AA129512 and AAM24494 to AAM24523 represent
 CC nucleotide and amino acid sequences given in the exemplification of the
 CC present invention
 XX
 SQ Sequence 158 AA;
 Query Match 100.0%; Score 878; DB 4; Length 158;
 Best Local Similarity 100.0%; Pred. No. 6.8e-83;
 Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASRSMRLLLLSCLAKTGVLDIMRPSCAPGWFYHKSNCYGFRLKLRMSDALEECQS 60
 DB 1 MASRSMRLLLLSCLAKTGVLDIMRPSCAPGWFYHKSNCYGFRLKLRMSDALEECQS 60
 QY 61 YGNGAHLASIIISLKEASTIAEYISGYORSQPIWIGLHPQKRQOMWIDGAMLYRSWSG 120
 DB 61 YGNGAHLASIIISLKEASTIAEYISGYORSQPIWIGLHPQKRQOMWIDGAMLYRSWSG 120
 QY 121 KSMGNGKHCAMSSNNNFLTWSNCKRQHFLLCKYRP 158
 DB 121 KSMGNGKHCAMSSNNNFLTWSNCKRQHFLLCKYRP 158
 RESULT 8
 AAM24520
 ID AAM24520 standard; protein; 158 AA.
 AC AAM24520;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE C880P similar amino acid sequence (GENESSEQ W37929).
 XX
 KW Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
 KW gene therapy; vaccine; colonic cancer.

XX OS Homo sapiens.
 XX PN MO200149716-A2.
 XX PD 12-JUL-2001.
 XX PF 29-DEC-2000; 2000MO-US0355596.
 XX PR 30-DEC-1999; 99US-00476296.
 XX PR 10-JAN-2000; 2000US-00480321.
 XX PR 15-FEB-2000; 2000US-00504629.
 XX PR 06-MAR-2000; 2000US-00519444.
 XX PR 19-MAY-2000; 2000US-00575251.
 XX PR 29-JUN-2000; 2000US-00609448.
 XX PR 28-AUG-2000; 2000US-00649811.
 XX PA (CORI-) CORIXA CORP.
 XX PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stork JA;
 XX PI King GE, Wang T, Jiang Y;
 XX DR WPI; 2001-441847/47.
 XX PT Colon tumor associated proteins and nucleic acids useful for the
 XX PT prevention, diagnosis and treatment of colonic cancer.
 XX PS Claim 2; Page 469; 472pp; English.
 XX CC The present invention describes colon tumour associated proteins (I) and
 XX CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 XX CC (I) and (II) can be used in gene therapy, diagnosis and vaccine production. (I) and
 XX CC (II) may be used in the prevention, diagnosis and treatment of diseases
 XX CC associated with inappropriate colon tumour associated protein (TCAP)
 XX CC expression, such as colonic cancer. For example, (I) and (II) may be used
 XX CC to treat disorders associated with decreased expression by rectifying
 XX CC mutations or deletions in a patient's genome that affect the activity of
 XX CC TCAPs by expressing inactive proteins or to supplement the patients own
 XX CC production of them. Additionally, (II) may be used to produce the TCAP
 XX CC proteins, by inserting the nucleic acids into a host cell culturing the
 XX CC cell to express the protein. (II) and its complementary sequences may
 XX CC also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
 XX CC and hybridisation assays to detect and quantitate the presence of similar
 XX CC nucleic acids in samples, and therefore which patients may be in need of
 XX CC restorative therapy. (I) may also be used as antigens in the production
 XX CC of antibodies against TCAPs and in assays to identify modulators of TCAP
 XX CC expression and activity. Anti-(I) antibodies and antagonists may also be
 XX CC used to down regulate TCAP expression and activity. The anti-(I)
 XX CC antibodies may also be used as diagnostic agents for detecting the
 XX CC presence of TCAPs in samples (e.g. by enzyme linked immunosorbent assay
 XX CC (ELISA)). AA128460 to AA129512 and AA24494 to AA24523 represent
 XX CC nucleotide and amino acid sequences given in the exemplification of the
 XX CC present invention
 XX SQ Sequence 158 AA;
 QY Query Match 100.0%; Score 878; DB 4; Length 158;
 Db Best Local Similarity 100.0%; Pred. No. 6.8e-83;
 Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASSMRLLLLSCLAKTGVGDIIMRPSCAPGWFYHNSNCYGFRKLRNMSDALEBQCS 60
 Db 1 MASRMRLLLLLSCLAKTGVGDIIMRPSCAPGWFYHNSNCYGFRKLRNMSDALEBQCS 60
 QY 61 YGNCAHLASIIISLKASTIAEYISGYORSQPIWIGLHPQKQOQWIDGMYLYRSWSG 120
 Db 61 YGNCAHLASIIISLKASTIAEYISGYORSQPIWIGLHPQKQOQWIDGMYLYRSWSG 120
 QY 121 KSMGNNKICAEWSSNNNFLTWSNNECNRRQHFLECKYRP 158
 Db 121 KSMGNNKICAEWSSNNNFLTWSNNECNRRQHFLECKYRP 158

RESULT 9
 ID AAM24521
 AC AAM24521; standard; protein, 158 AA.
 AC AAM24521;
 DT 12-OCT-2001 (first entry)
 XX DE C880P similar amino acid sequence (GENESQ W84274).
 XX KW Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
 XX KW gene therapy; vaccine; colonic cancer.
 XX OS Homo sapiens.
 XX PN MO200149716-A2.
 XX PD 12-JUL-2001.
 XX PF 29-DEC-2000; 2000MO-US0355596.
 XX PR 30-DEC-1999; 99US-00476296.
 XX PR 10-JAN-2000; 2000US-00480321.
 XX PR 15-FEB-2000; 2000US-00504629.
 XX PR 06-MAR-2000; 2000US-00519444.
 XX PR 19-MAY-2000; 2000US-00575251.
 XX PR 29-JUN-2000; 2000US-00609448.
 XX PR 28-AUG-2000; 2000US-00649811.
 XX PA (CORI-) CORIXA CORP.
 XX PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stork JA;
 XX PI King GE, Wang T, Jiang Y;
 XX DR WPI; 2001-441847/47.
 XX PT Colon tumor associated proteins and nucleic acids useful for the
 XX PT prevention, diagnosis and treatment of colonic cancer.
 XX PS Claim 2; Page 469; 472pp; English.
 XX CC The present invention describes colon tumour associated proteins (I) and
 XX CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 XX CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 XX CC (II) may be used in the prevention, diagnosis and treatment of diseases
 XX CC associated with inappropriate colon tumour associated protein (TCAP)
 XX CC expression, such as colonic cancer. For example, (I) and (II) may be used
 XX CC to treat disorders associated with decreased expression by rectifying
 XX CC mutations or deletions in a patient's genome that affect the activity of
 XX CC TCAPs by expressing inactive proteins or to supplement the patients own
 XX CC production of them. Additionally, (II) may be used to produce the TCAP
 XX CC proteins, by inserting the nucleic acids into a host cell culturing the
 XX CC cell to express the protein. (II) and its complementary sequences may
 XX CC also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
 XX CC and hybridisation assays to detect and quantitate the presence of similar
 XX CC nucleic acids in samples, and therefore which patients may be in need of
 XX CC restorative therapy. (I) may also be used as antigens in the production
 XX CC of antibodies against TCAPs and in assays to identify modulators of TCAP
 XX CC expression and activity. Anti-(I) antibodies and antagonists may also be
 XX CC used to down regulate TCAP expression and activity. The anti-(I)
 XX CC antibodies may also be used as diagnostic agents for detecting the
 XX CC presence of TCAPs in samples (e.g. by enzyme linked immunosorbent assay
 XX CC (ELISA)). AA128460 to AA129512 and AA24494 to AA24523 represent
 XX CC nucleotide and amino acid sequences given in the exemplification of the
 XX CC present invention
 XX SQ Sequence 158 AA;
 QY Query Match 100.0%; Score 878; DB 4; Length 158;
 Db Best Local Similarity 100.0%; Pred. No. 6.8e-83;
 Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASRMRLLLLLSCLAKTGVGDIIMRPSCAPGWFYHNSNCYGFRKLRNMSDALEBQCS 60

```

Db      1 MASRRMLLLLSCLATGVLGDIIMRSCAPGWFYHKSNCYGYFRKLRNMSDALEECQS 60
Qy      61 YGNGAHLASTLSTKEASTIAEYISGYORSOPITWIGLHDPQRQOWIDGAMLYLRSMG 120
Db      61 YGNGAHLASTLSTKEASTIAEYISGYORSOPITWIGLHDPQRQOWIDGAMLYLRSMG 120
Qy      121 KSMGNKHCAMSSNNNFLTWSNCKNRQHFCKYRP 158
Db      121 KSMGNKHCAMSSNNNFLTWSNCKNRQHFCKYRP 158

RESULT 10
AAM24517
ID      AAM24517 standard; protein; 158 AA.
XX
XX      AAM24517;
AC
XX      12-OCT-2001 (first entry)
DT
XX      CS1-152 clone predicted amino acid sequence.
DE
XX      Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
KM      gene therapy; vaccine; colonic cancer.
XX
XX      Homo sapiens.
OS
XX      MO200149716-A2.
FN
XX      12-JUL-2001.
PD
XX      29-DEC-2000; 2000WO-US035596.
PE
XX      30-DEC-1999; 99US-00476296.
PR      10-JAN-2000; 2000US-00480321.
PR      15-FEB-2000; 2000US-00504629.
PR      06-MAR-2000; 2000US-00519444.
PR      19-MAY-2000; 2000US-00575251.
PR      29-JUN-2000; 2000US-00609448.
PR      28-AUG-2000; 2000US-00649811.
XX
XX      (CORI-) CORIXA CORP.
PA
XX      Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI      King GE, Wang T, Jiang Y;
XX
XX      WPI; 2001-441847/47.
PT      Colon tumor associated proteins and nucleic acids useful for the
PS      prevention, diagnosis and treatment of colonic cancer.
XX
XX      Claim 2; Page 463; 472pp; English.
CC      The present invention describes colon tumour associated proteins (I) and
CC      the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC      (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC      (II) may be used in the prevention, diagnosis and treatment of diseases
CC      associated with inappropriate colon tumour associated protein (TCAP)
CC      expression, such as colonic cancer. For example, (I) and (II) may be used
CC      to treat disorders associated with decreased expression by rectifying
CC      mutations or deletions in a patient's genome that affect the activity of
CC      TCAPs by expressing inactive proteins or to supplement the patient's own
CC      production of them. Additionally, (II) may be used to produce the TCAP
CC      proteins, by inserting the nucleic acids into a host cell culturing the
CC      cell to express the protein. (II) and its complementary sequences may
CC      also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
CC      and hybridisation assays to detect and quantitate the presence of similar
CC      nucleic acids in samples, and therefore which patients may be in need of
CC      restorative therapy. (I) may also be used as antigens in the production
CC      of antibodies against TCAPs and in assays to identify modulators of TCAP
CC      expression and activity. Anti-(I) antibodies and antagonists may also be
CC      used to down regulate TCAP expression and activity. The anti-(I)
CC      antibodies may also be used as diagnostic agents for detecting the

```

```

CC      presence of TCAPs in samples (e.g. by enzyme linked immunosorbant assay
CC      (ELISA)). AA128460 to AA129512 and AAM24494 to AAM24523 represent
CC      nucleotide and amino acid sequences given in the exemplification of the
CC      present invention
XX
XX      Sequence 158 AA;
SQ
Query Match      100.0%; Score 878; DB 4; Length 158;
Best Local Similarity 100.0%; Pred. No. 6, 8e-83;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MASRRMLLLLSCLATGVLGDIIMRSCAPGWFYHKSNCYGYFRKLRNMSDALEECQS 60
Db      1 MASRRMLLLLSCLATGVLGDIIMRSCAPGWFYHKSNCYGYFRKLRNMSDALEECQS 60
Qy      61 YGNGAHLASTLSTKEASTIAEYISGYORSOPITWIGLHDPQRQOWIDGAMLYLRSMG 120
Db      61 YGNGAHLASTLSTKEASTIAEYISGYORSOPITWIGLHDPQRQOWIDGAMLYLRSMG 120
Qy      121 KSMGNKHCAMSSNNNFLTWSNCKNRQHFCKYRP 158
Db      121 KSMGNKHCAMSSNNNFLTWSNCKNRQHFCKYRP 158

RESULT 11
AAM24518
ID      AAM24518 standard; protein; 158 AA.
XX
XX      AAM24518;
AC
XX      12-OCT-2001 (first entry)
DT
XX      C880P similar amino acid sequence (GENEBQ W12691).
DE
XX      Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
KM      gene therapy; vaccine; colonic cancer.
XX
XX      Homo sapiens.
OS
XX      MO200149716-A2.
FN
XX      12-JUL-2001.
PD
XX      29-DEC-2000; 2000WO-US035596.
PE
XX      30-DEC-1999; 99US-00476296.
PR      10-JAN-2000; 2000US-00480321.
PR      15-FEB-2000; 2000US-00504629.
PR      06-MAR-2000; 2000US-00519444.
PR      19-MAY-2000; 2000US-00575251.
PR      29-JUN-2000; 2000US-00609448.
PR      28-AUG-2000; 2000US-00649811.
XX
XX      (CORI-) CORIXA CORP.
PA
XX      Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI      King GE, Wang T, Jiang Y;
XX
XX      WPI; 2001-441847/47.
PT      Colon tumor associated proteins and nucleic acids useful for the
PS      prevention, diagnosis and treatment of colonic cancer.
XX
XX      Claim 2; Page 467; 472pp; English.
CC      The present invention describes colon tumour associated proteins (I) and
CC      the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC      (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC      (II) may be used in the prevention, diagnosis and treatment of diseases
CC      associated with inappropriate colon tumour associated protein (TCAP)
CC      expression, such as colonic cancer. For example, (I) and (II) may be used
CC      to treat disorders associated with decreased expression by rectifying
CC      mutations or deletions in a patient's genome that affect the activity of

```

PT of a Reg Like Protein or the presence or quantity of a nucleic acid

CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be

CC used in antisense therapy. An antibody immunoreactive with a polypeptide
CC encoded by (1) is useful for detecting cancer in a patient sample, and
CC for detecting the presence or absence of a polynucleotide encoded by a
CC nucleic acid which hybridizes to (1) in a cell. A probe/primer derived
CC from (1) can be used for determining the presence of a nucleic acid which
CC hybridizes to (1), and for determining the phenotype of cells in a sample
CC of cells from a patient. (1) is useful for determining the presence of
CC colon cancer in a cell or tissue type, for determining the presence or
CC state of other type of cancer, in antisense therapy, to generate
CC macroarrays on a solid surface, to identify a chromosome on which the
CC corresponding gene resides, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (1) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists
CC
XX Sequence 158 AA;

Query Match 100.0%; Score 878; DB 5; Length 158;
Best Local Similarity 100.0%; Pred. No. 6.8e-83;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRSMRLLLLSCLAKTGVLDIIMRSPCAPGWFYHNSCYGFRKLRNMSDALEBQS 60
DB 1 MASRSMRLLLLSCLAKTGVLDIIMRSPCAPGWFYHNSCYGFRKLRNMSDALEBQS 60
QY 61 YNGNAHLASISLKEASTIAEYISGYORSQPIWIGLHPQKRQOWIDGAMYLRYRMSG 120
DB 61 YNGNAHLASISLKEASTIAEYISGYORSQPIWIGLHPQKRQOWIDGAMYLRYRMSG 120
QY 121 KSMGNKHCAMSSNNNFLTWSNECNKRQHFCKYRP 158
DB 121 KSMGNKHCAMSSNNNFLTWSNECNKRQHFCKYRP 158

RESULT 14
ABRS551
ID ABRS551 standard; protein; 158 AA.

AC ABRS551;

DT 09-JUN-2003 (first entry)

DE Human cancer related protein SEQ ID NO:208.

KW Human; cancer; diagnosis; screening; modulator; leukemia; ischemia;
KW heart disease; atherosclerosis; endometriosits.

OS Homo sapiens.

PN WO2003025138-A2.

PD 27-MAR-2003.

PF 17-SEP-2002; 2002WO-US029560.

PR 17-SEP-2001; 2001US-0323469P.

PR 20-SEP-2001; 2001US-0323887P.

PR 13-NOV-2001; 2001US-0350666P.

PR 08-FEB-2002; 2002US-0355145P.

PR 08-FEB-2002; 2002US-0355257P.

PR 12-APR-2002; 2002US-0372246P.

PA (EOSB-) EOS BIOTECHNOLOGY INC.

PI Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;

PI Zlotnick A;

DR WPI; 2003-354600/33.

DR N-PSDB; ACC72672.

XX New genes that are up-regulated or down-regulated in cancers, useful as
XX markers for diagnosing e.g. cancer, ischemia or heart diseases, or as
XX therapeutic targets for screening drugs for treating these diseases.

PS Claim 12; Page 737; 767pp; English.

XX The present invention describes an isolated nucleic acid molecule, which
CC comprises the sequence of any of the genes that are up-regulated or down-
CC regulated in specific cancers (e.g. about 1031 genes up-regulated in
CC acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer
CC related gene nucleotide sequences which encode the proteins given in
CC ABRS551 to ABRS709. Also described: (1) determining the presence or
CC absence of a pathological cell in a patient; (2) an expression vector
CC comprising a nucleic acid molecule described above; (3) a host cell
CC comprising the vector; (4) an isolated polypeptide, which is encoded by
CC the nucleic acid; (5) an antibody that specifically binds the polypeptide
CC of (4); (6) specifically targeting a compound to a pathological cell in a
CC patient by administering to the patient the antibody above; and (7) a
CC drug screening assay. The nucleic acid is useful as diagnostic markers or
CC therapeutic targets. In particular, the nucleic acid is useful for
CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,
CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,
CC pancreas, prostate, skin and uterus), wounds, ischemia, heart diseases,
CC atherosclerosis and endometriosits. The nucleic acid is also useful in
CC drug screening, particularly for identifying agents for treating these
CC pathologies
CC
XX

SQ Sequence 158 AA;

Query Match 100.0%; Score 878; DB 6; Length 158;
Best Local Similarity 100.0%; Pred. No. 6.8e-83;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRSMRLLLLSCLAKTGVLDIIMRSPCAPGWFYHNSCYGFRKLRNMSDALEBQS 60
DB 1 MASRSMRLLLLSCLAKTGVLDIIMRSPCAPGWFYHNSCYGFRKLRNMSDALEBQS 60

QY 61 YNGNAHLASISLKEASTIAEYISGYORSQPIWIGLHPQKRQOWIDGAMYLRYRMSG 120
DB 61 YNGNAHLASISLKEASTIAEYISGYORSQPIWIGLHPQKRQOWIDGAMYLRYRMSG 120

QY 121 KSMGNKHCAMSSNNNFLTWSNECNKRQHFCKYRP 158
DB 121 KSMGNKHCAMSSNNNFLTWSNECNKRQHFCKYRP 158

RESULT 15
ABPS6022
ID ABPS6022 standard; protein; 158 AA.

AC ABPS6022;

DT 26-FEB-2003 (first entry)

DE Human REG-like protein (RELIP) SEQ ID NO:2.

KW Human; REG-like protein; RELIP; immunoglobulin derived protein; Ig;
KW immunoglobulin; cytosolic; Ig agonist; immunoglobulin agonist; cancer;
KW protein therapy; RELIP human Ig derived protein; chromosome 1p12-13.1.

OS Homo sapiens.

PN WO200274916-A2.

PD 26-SEP-2002.

PF 14-MAR-2002; 2002WO-US007945.

PF 16-MAR-2001; 2001US-0276305P.

PA (CENZ) CENTOCOR INC.

PI Key Location/Qualifiers

PI Peptide 1..26

PI Protein 27..158

PI label= RELIP

XX Heiskala M;

DR WPI; 2003-103204/09.

DR N-PSDB; AB821635.

PT New isolated REG-like protein (RELp) human immunoglobulin derived protein
PT or specified portion or variant, useful for preventing or treating a RELp
PT protein mediated condition or malignant condition, e.g. cancer.

PS Claim 1; Fig 2; 101pp; English.

CC The present sequence represents a new isolated REG-like protein (RELp)
CC human immunoglobulin (Ig) derived protein. RELp comprises: (a) a human
CC variable and constant region; or (b) an isolated human Ig derived protein
CC or specified portion or variant encoded by a nucleic acid. RELp has
CC cytostatic activity and can be used as an Ig agonist and in protein
CC therapy. The RELp human Ig derived protein or a specified portion or
CC variant can be used for preventing or treating a RELp protein mediated
CC condition, malignant condition or disease condition, e.g. cancer. The
CC nucleic acids can be used in producing RELp Ig derived protein. The human
CC RELp protein of the present invention is located to chromosome 1p12-13.1

CC Sequence 158 AA;

Query Match 100.0%; Score 878; DB 6; Length 158;
Best Local Similarity 100.0%; Pred. No. 6.8e-83;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRSMTLLLSLCTAKTGVLDIIMRPSCAPGWFYHKSNCYGYFRKLRNMSDALEECOS 60
DB 1 MASRSMTLLLSLCTAKTGVLDIIMRPSCAPGWFYHKSNCYGYFRKLRNMSDALEECOS 60
QY 61 YGNGAHLASISLKEASTIAEYISGYORSQPIWIGLHDPQKROQWQWIDGAMTYLYRSWSG 120
DB 61 YGNGAHLASISLKEASTIAEYISGYORSQPIWIGLHDPQKROQWQWIDGAMTYLYRSWSG 120
QY 121 KSMGNGKCAEMSSNNNFLTWSNNECNKROHFLCKYRP 158
DB 121 KSMGNGKCAEMSSNNNFLTWSNNECNKROHFLCKYRP 158

Search completed: February 11, 2005, 23:29:37
Job time : 169 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2005, 23:16:55 ; Search time 177 Seconds

(without alignments)
457.110 Million cell updates/sec

Title: US-09-525-041-2

Sequence: 1 MASRSMRLLLLSCLAKGV.....LTWSNECKRQHFLCKTRP 158

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	878	100.0	158	2	09BYZ8
2	600.5	68.4	157	2	09BDG5
3	594.5	67.7	157	2	09BD58
4	593.5	67.6	157	2	068AX7
5	544	62.0	113	2	08NER7
6	438.5	49.9	160	2	075X75
7	307.5	35.0	134	2	08NER6
8	260.5	29.7	132	1	STR1_STRCA
9	258.5	29.4	132	1	ACAL_ANSAN
10	254.5	29.0	165	1	LIT1_MOUSE
11	254.5	29.0	165	1	LITR_RAT
12	247	28.1	142	1	STR2_STRCA
13	246.5	28.1	174	1	PAP3_MOUSE
14	245	27.9	164	2	06TR56
15	243.5	27.7	158	2	06QX33
16	242	27.6	175	1	PAP2_MOUSE
17	241.5	27.5	142	1	OC17_CHICK
18	240	27.3	166	1	LITR_HUMAN
19	237.5	27.0	158	2	071R01
20	237	27.0	135	1	LECG_BOTUR
21	236.5	26.9	173	1	LIT2_MOUSE
22	236	26.9	126	2	08CEP9
23	235	26.8	126	2	08CEP9
24	234	26.7	135	1	LECG_BITAR
25	233	26.5	135	1	LECG_CROAT
26	231	26.3	174	1	PAP3_RAT
27	231	26.3	175	1	LITR_BOVIN
28	230.5	26.3	154	2	08BIV9
29	229	26.1	135	1	LECG_LACST
30	228.5	26.0	174	1	PAP2_RAT
31	227.5	25.9	146	2	09CVF4

ALIGNMENTS

RESULT 1	PRELIMINARY	PRT	158 AA.
09BYZ8			
AC	09BYZ8;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	25-OCT-2004 (TrEMBLrel. 28, Last annotation update)		
DE	Regenerating gene type IV precursor (REG-like protein) (Regenerating		
DE	islet-derived family, member 4) (Gastrointestinal secretory protein		
DE	GISP)		
GN	Name=REG4;		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=21210973; PubMed=11311942; DOI=10.1016/S0167-4781(00)00284-0;		
RA	Hartup J.C., Zhang H., Bonaldi M.F., Soares M.B., Dieckgraber B.K.;		
RT	"Isolation and characterization of a cDNA encoding a novel member of		
RT	the human regenerating protein family: Reg IV(1)."		
RL	Biochim. Biophys. Acta 1518:287-293(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Kamranian M., Heiskala K., Heiskala M., Anderson L.C.;		
RA	Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Colon;		
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;		
RA	Strassberg R.L., Feingold E.A., Grouse L.H., Derge J.G.;		
RA	Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.;		
RA	Altschul S.F., Zeeberg B., Burow K.H., Schaefer C.F., Bat N.K.;		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.;		
RA	Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.;		
RA	Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.;		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.;		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.;		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.;		
RA	Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.;		
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.;		
RA	Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.;		
RA	Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.;		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.;		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.;		
RA	Krzyniaki M.I., Skalska U., Smalits D.E., Scherch A., Schein J.E.;		
RA	Jones S.J., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		

32	227	25.9	158	2	09OW17
33	227	25.9	158	2	09OW18
34	225.5	25.7	146	2	06X5S1
35	225.5	25.7	146	2	06X5S4
36	225.5	25.7	166	1	LITR_HUMAN
37	222.5	25.3	172	1	LECG_PLEMA
38	222	25.3	157	2	06T7B5
39	221.5	25.2	155	2	08TIV8
40	221	25.2	175	1	PAP1_HUMAN
41	219.5	25.0	148	2	06X5S3
42	219.5	25.0	148	2	06X5S7
43	219.5	25.0	148	2	06X5S9
44	218.5	25.0	148	2	06X5T1
45	218.5	24.9	154	2	07T2Q0

09OW17	bungarus fa
09OW18	bungarus fa
06X5S1	echis pyram
06X5S4	echis carin
P05451	homo sapien
002968	pleurodeles
06T7B5	bitis gabon
08TIV8	agkistrodon
006141	homo sapien
06X5S3	echis pyram
06X5S7	echis ocell
06X5S9	echis carin
06X5T1	bitis ariet
07T2Q0	echis mult

RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Straubeberg R.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Lin W.-C.;
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY007243; AAG02562.1; -
 DR EMBL; AF345934; AAK59869.1; -
 DR EMBL; AY126670; AAM95598.1; -
 DR EMBL; BC017089; AAH17089.1; -
 DR EMBL; AF254415; AAK48435.1; -
 DR HSSP; P22030; I1IK.
 DR Genew; HGNC:22977; REG4.
 DR GO; GO:0005529; F: sugar binding; IEA.
 DR InterPro; IPR01304; Lectin_C.
 DR InterPro; IPR03990; Pancreatic_ac.
 DR Pfam; PF00059; Lectin_C_1.
 DR PRINTS; PR01504; PNCEARITSAF.
 DR SMART; SM00034; CLECT_1.
 DR PROSITE; PS00615; C_TYPE_LECTIN_1; UNKNOWN_1.
 DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
 KW Signal.
 FT SIGNAL.
 SQ SEQUENCE 158 AA; 18230 MW; 7308849CBBD6E93E CRC64;
 Query Match 100.0%; Score 878; DB 2; Length 158;
 Best Local Similarity 100.0%; Pred. No. 2,5e-78;
 Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRSMRLILSLCTAGTGTGDIIMRPSGAPGWFYKSNCGYPRKLRNMSDALEECQS 60
 DB 1 MASRSMRLILSLCTAGTGTGDIIMRPSGAPGWFYKSNCGYPRKLRNMSDALEECQS 60

QY 61 YGNGHHLASITSLKASTAETISGYORSOPITGLHPQKQOWIDGAMTYLYRMSG 120
 DB 61 YGNGHHLASITSLKASTAETISGYORSOPITGLHPQKQOWIDGAMTYLYRMSG 120

QY 121 KSMGNGKCAEWSNNFLLTSSNCCNKRQHFLLCKYRP 158
 DB 121 KSMGNGKCAEWSNNFLLTSSNCCNKRQHFLLCKYRP 158

QY 121 KSMGNGKCAEWSNNFLLTSSNCCNKRQHFLLCKYRP 158
 DB 121 KSMGNGKCAEWSNNFLLTSSNCCNKRQHFLLCKYRP 158

RESULT 2
 Q9D8G5 PRELIMINARY; PRT; 157 AA.
 ID Q9D8G5;
 AC Q9D8G5;
 DT 01-UN-2001 (TREMBlrel. 17, Created)
 DT 01-UN-2001 (TREMBlrel. 17, Last sequence update)
 DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
 DE Mus musculus adult male small intestine cDNA, RIKEN full-length
 DE enriched library, clone:2010002L15 product:REGENERATING GENE TYPE IV,
 DE full insert sequence (Reg4 protein).
 GN Name=Reg4;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Small intestine;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Mech. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Small intestine;
 RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;

RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Small intestine;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Small intestine;
 RC MEDLINE=204993174; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Small intestine;
 RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Alzawa K., Nagaoaka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kitasuna T., Tashiro H., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaishogi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanaki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Small intestine;
 RA Adachi T., Alzawa K., Akihara S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukuishi Y., Furuno M.,
 RA Hasegawa T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Kono H., Kouda M., Koyu S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numasaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RC MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schutler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshitsuki S., Carninci P., Prange C.,
 RA Rana S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butcherfield Y.S.,
 RA Krzywinski M.I., Skalek U., Smallos D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Warr M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

```

RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK008049; BAB25429.1; -
DR EMBL; BC019465; AAH19465.1; -
DR HSSP; Q06141; IUV0.
DR MGD; MGI:1914959; Reg4.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR003990; Pancratic_ac.
DR Pfam; PF00059; Lectin_C.1.
DR PRINTS; PR01504; PNCEATITSP.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PSS0041; C_TYPE_LLECTIN_2; 1.
SQ SEQUENCE 157 AA; 18398 MW; F3981722BBD83968 CRC64;

Query Match 68.4%; Score 600.5; DB 2; Length 157;
Best Local Similarity 66.2%; Pred. No. 4,8e-51;
Matches 104; Conservative 25; Mismatches 27; Indels 1; Gaps 1;

QY 1 MASRSRRLLLLSCTAKTGVLDIMRPSCAPGWFYHNSGCGYRKJLNMSDALEECOS 60
DB 1 MASRGRRLLLLSWVAGPEVLSD-ILRPSCAPGWFYHNSGCGYRKJLNMSHALEECOS 59
QY 61 YNGNHLASITSLKASTIAETISGYORSOPITWIGLHDPQKQOWIDGAMTYLRWSG 120
DB 60 YNGSHLASVNLQKQASVISKYITGQRLPWIGLHDPQKQOWIDGSHNLYRRWNP 119
QY 121 KSMGNKCAEMSSNNFLTWSNCKRQHFLLCYR 157
DB 120 RTKSEARHCAMNPKDKFLTNKNGCANRQHFLLCYR 156

RESULT 3
Q9D858 PRELIMINARY; PRT; 157 AA.
ID Q9D858
AC Q9D858;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male small intestine cDNA, RIKEN full-length
DE enriched library, clone:2010204K21 product:REGENERATING GENE TYPE IV,
DE full insert sequence.
GN Name=Reg4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small Intestine;
RA MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RA "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44 (1999);
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small Intestine;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small Intestine;
RA The FANTOM Consortium;
RA "Analysis of the mouse transcriptome based on functional annotation of
RA 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
[4]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small Intestine;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630 (2000).
[5]
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small Intestine;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Atzawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama U., Nishi K., Kitsuana T., Teshiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771 (2000).
[6]
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small Intestine;
RA Adachi J., Atzawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Furukoshi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imocani K., Ishii Y., Itoh M., Izawa M., Kaekawa T., Kato H.,
RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ono M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shiba K., Shiga K., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Teijima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK008438; BAB25669.1; -
DR HSSP; Q06141; IUV0.
DR MGD; MGI:1914959; Reg4.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR003990; Pancratic_ac.
DR Pfam; PF00059; Lectin_C.1.
DR PRINTS; PR01504; PNCEATITSP.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PSS0041; C_TYPE_LLECTIN_2; 1.
SQ SEQUENCE 157 AA; 18474 MW; FD96F36CFB989368 CRC64;

Query Match 67.7%; Score 594.5; DB 2; Length 157;
Best Local Similarity 65.6%; Pred. No. 1.9e-50;
Matches 103; Conservative 25; Mismatches 28; Indels 1; Gaps 1;

QY 1 MASRSRRLLLLSCTAKTGVLDIMRPSCAPGWFYHNSGCGYRKJLNMSDALEECOS 60
DB 1 MAYKGVRLLLLSWVAGPEVLSD-ILRPSCAPGWFYHNSGCGYRKJLNMSHALEECOS 59
QY 61 YNGNHLASITSLKASTIAETISGYORSOPITWIGLHDPQKQOWIDGAMTYLRWSG 120
DB 60 YNGSHLASVNLQKQASVISKYITGQRLPWIGLHDPQKQOWIDGSHNLYRRWNP 119
QY 121 KSMGNKCAEMSSNNFLTWSNCKRQHFLLCYR 157
DB 120 RTKSEARHCAMNPKDKFLTNKNGCANRQHFLLCYR 156

RESULT 4
Q68AX7 PRELIMINARY; PRT; 157 AA.
ID Q68AX7
AC Q68AX7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)

```

```
DE Regenerating islet-derived family member 4.
GN Name=Reg4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mistary, TISSUE=ileum;
RA Naitkawa K., Murakami K., Fukushina M., Kiyama H.;
RT "Differential regulation of Reg family member expression after
RT peripheral nerve injury.";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB164049; BAD38673.1; -.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR003990; Pancreatls_ac.
DR Pfam; PF00059; Lectin_C.1.
DR PRINTS; PR01504; PNCREATITSAP.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
SQ SEQUENCE 157 AA; 18265 MW; 55B129FB10BA4D1D CRC64;

Query Match 67.6%; Score 593.5; DB 2; Length 157;
Best Local Similarity 66.5%; Pred. No. 2.3e-50;
Matches 105; Conservative 21; Mismatches 31; Indels 1; Gaps 1;

Qy 1 MASRMRLLLSLCLAKTGVIGDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEECOS 60
Db 1 MASKKVRLLLSLWAGPEVLSD-ILRPSCASGMWYNRSHCHGFRKLRNMSHALEECOS 59
Qy 61 YGNGAHLASILSLKEASTIAEYISGYORSOPITWIGLHPDKRQOMQWIDGAMYLRSMSG 120
Db 60 YGNGSHLASLVNPKKASVISKTIYQKRSPLPWIGLHPQKNASQWIDGSTNQVRPMS 119
Qy 121 KSMGNGKCAEWSNNNFLTMSNCECNKRQHFLECYRP 158
Db 120 RTKSEARHCTEMNPDKFELTWKNGCTKQHFLECYRP 157

RESULT 5
QENER7 PRELIMINARY; PRT; 113 AA.
AC Q8NER7;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE REG-like protein splice variant 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kamatani M., Heiskala K., Heiskala M., Andersson L.C.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY126671; AAM95599.1; -.
DR HSSP; P21963; IJZN.
DR GO; GO:0005529; F:sugar binding; IEA.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR003990; Pancreatls_ac.
DR Pfam; PF00059; Lectin_C.1.
DR PRINTS; PR01504; PNCREATITSAP.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
SQ SEQUENCE 113 AA; 12832 MW; A2E9DFA729C78DA CRC64;

Query Match 62.0%; Score 544; DB 2; Length 113;
Best Local Similarity 99.0%; Pred. No. 1.2e-45;
Matches 101; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MASRMRLLLSLCLAKTGVIGDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEECOS 60
Db 1 MASRMRLLLSLCLAKTGVIGDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEECOS 60
```

```
Qy 61 YGNGAHLASILSLKEASTIAEYISGYORSOPITWIGLHPDKR 102
Db 61 YGNGAHLASILSLKEASTIAEYISGYORSOPITWIGLHPDKR 102

RESULT 6
Q8275 PRELIMINARY; PRT; 160 AA.
ID Q8275;
AC Q8275;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE MG64513 protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=whole;
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Bromanstein M.J., Usdin T.B., Toshiyuki S., Carantini P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman W., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Boulford G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grilms J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=whole;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT Initiative."
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=whole;
RA Klein S., Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053817; AAHS3817.1; -.
DR HSSP; P22897; IREG.
DR GO; GO:0005529; F:sugar binding; IEA.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; Lectin_C.1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
SQ SEQUENCE 160 AA; 18330 MW; 6A5502F24689179A CRC64;

Query Match 49.9%; Score 438.5; DB 2; Length 160;
Best Local Similarity 48.7%; Pred. No. 4.1e-35;
Matches 75; Conservative 35; Mismatches 41; Indels 3; Gaps 2;

Qy 8 LLLLSLCLAKTGVIGDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEECOSYGNGAHL 67
Db 8 LLLLSLCLAKTGVIGDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEECOSYGNGAHL 67
```

```

DB 7 LLLPLGALAVSNVLEAAYRSSCPNGMFPYKANCYGFYRPLSMABEAYDCQAYGHAMU 66
QY 68 ASTLSLKEASTAEVYSGVRSOPITVGLHPDROKQOMQIDAMLYLSW-SGKSMGN 126
DB 67 ASTLDSAEADVASHSAVQKNPVMIGLHDPONRMRKMDSMYNYSWLAGDPDNYN 126
QY 127 --KHCAEMSSNNFLTWSSNECNKROHFLCKYRP 158
DB 127 SAEYCELSCKEFGVAKMNSNCKEYVQYCKYRP 160

RESULT 7
Q8MER6 PRELIMINARY; PRT; 134 AA.
ID Q8MER6;
AC Q8MER6;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DE REG-like protein splice variant 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Karamainen M., Heiskala K., Heiskala M., Anderson L.C.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; A126672; AM93960-1;
SQ SEQUENCE 134 AA; 14993 MW; E0E5AD9B6A53EB5 CRC64;

Query Match 35.0%; Score 307.5; DB 2; Length 134;
Best local Similarity 55.9%; Pred. No. 2.6e-22;
Matches 62; Conservative 9; Mismatches 9; Indels 31; Gaps 3;

QY 1 MASRSNRLLLLSCLAKTGYLGDIMRPSCAPMFPYKNSCYGFKLRNWSDAEL---- 56
DB 1 MASRSNRLLLLSCLAKTGYLGDIMRPSCAPMFPYKNSCYGFKLRNWSDAELVRL 60
QY 57 -----ECOSYNGAHLASLSLKEASTAEVYSGVRSOPITV 94
DB 61 PAMFGLSRAKDQPEPQ-----ISFDSGSSV--LPGHYEKPLML 97

RESULT 8
STR1_STRCA STANDARD; PRT; 132 AA.
ID STR1_STRCA
AC P83514;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Struthio camelus (Ostrich).
OS Struthio camelus (Ostrich).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Struthioniformes; Struthionidae;
OC Struthio.
OX NCBI_TaxID=8801;
RN [1]
RP SEQUENCE, SUBCELLULAR LOCATION, AND MASS SPECTROMETRY.
RC TISSUE=Eggshell matrix;
RX PubMed=14726203; DOI=10.1016/j.bbapap.2003.09.006;
RA Mann K., Siedler F.;
RT "Ostrich (Struthio camelus) eggshell matrix contains two different C-
RT type lectin-like proteins. Isolation, amino acid sequence, and
RT posttranslational modifications."
RL Biochim. Biophys. Acta 1696:41-50(2004).
CC -1- SUBCELLULAR LOCATION: Eggshell matrix.
CC -1- MASS SPECTROMETRY: MW=15343.2; MW_ERR=4; METHOD=Electrospray;
CC RANGE=1-133; NOTE=Ref.1.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
DR InterPro; IPR002353; Antifreeze1.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR003990; Pancreatit_ac.
DR Pfam; PF00059; Pancreatit_ac; 1.

```

```

DR PRINTS; PR00356; ANTIFREEZE11.
DR PRINTS; PR01504; PNCREATITSAF.
DR SMART; SM00034; CLECT. 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Direct protein sequencing; Lectin.
FT DOMAIN 10 129 C-type lectin.
FT DISULFID 3 14 By similarity.
FT DISULFID 31 126 By similarity.
FT DISULFID 103 120 By similarity.
SQ SEQUENCE 132 AA; 15353 MW; F7BDIDF2990B2945 CRC64;

Query Match 29.7%; Score 260.5; DB 1; Length 132;
Best local Similarity 38.8%; Pred. No. 1.1e-17;
Matches 50; Conservative 20; Mismatches 54; Indels 5; Gaps 2;

QY 30 CAPMFPYKNSCYGFKLRNWSDAELFCOSYNGAHLASLSLKEASTAEVYSGY--- 86
DB 3 CPKGMIDFRNGCYGFYRELPMKRAEAWCRSIRAGHNLASIHITSEHRAIAKFIISQYHHG 62
QY 87 QRSOPITVGLHPDROKQOMQIDGAMLYLSRNSGKSMGNKCAEMSSNNFLTWSSNRC 146
DB 63 EEEEDVWIGLF--RMSVWAWIDGSKKHSALDDDDYPRKHCAYLDSSGFLSWDNDSC 120
QY 147 NKROHFLCK 155
DB 121 GERNAFICK 129

RESULT 9
ACAL_ANSAN STANDARD; PRT; 132 AA.
ID ACAL_ANSAN
AC P83300;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Ansoalcain.
OS Anser anser (Western graylag goose).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anser.
OX NCBI_TaxID=8844;
RN [1]
RP SEQUENCE, FUNCTION, SUBUNIT, SUBCELLULAR LOCATION, AND MASS
RP SPECTROMETRY.
RC TISSUE=Eggshell matrix;
RX MEDLINE=22439773; PubMed=12431998; DOI=10.1074/jbc.M201518200;
RA Lakshminarayana R., Valliyaveetil S., Rao V.S., Kini R.M.;
RT "Purification, characterization, and in vitro mineralization studies
RT of a novel goose eggshell matrix protein, ansocalcin."
RL J. Biol. Chem. 278:2928-2936(2003).
CC -1- FUNCTION: Induces spherical aggregates of calcite crystals in
CC vitro. Believed to play an active role in the eggshell
CC calcification.
CC -1- SUBUNIT: Homodimer or homotrimer.
CC -1- SUBCELLULAR LOCATION: Eggshell matrix. May be present in the shell
CC glands on the walls of oviduct and incorporated into the shell
CC structure during its formation.
CC -1- MASS SPECTROMETRY: MW=15342; METHOD=Electrospray; RANGE=1-132;
CC NOTE=Ref.1.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
DR HSP; P23806; L134.
DR InterPro; IPR002353; Antifreeze1.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR003990; Pancreatit_ac.
DR Pfam; PF00059; Lectin_C; 1.
DR PRINTS; PR00356; ANTIFREEZE11.
DR PRINTS; PR01504; PNCREATITSAF.
DR SMART; SM00034; CLECT. 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Direct protein sequencing; Lectin.
FT DOMAIN 1 132 C-type lectin.
FT DISULFID 3 14 By similarity.

```

```

FT DISULFID 31 128 By similarity.
FT DISULFID 103 120 By similarity.
SQ SEQUENCE 132 AA; 15347 MW; 36CE42BA57226B7 CRC64;

Query Match
Beet Local Similarity 34.3%; Pred. No. 1.7e-17;
Matches 46; Conservative 25; Mismatches 50; Indels 13; Gaps 3;

QY 30 CAPGFYHNSCYGFRKLRMSDAELPGQSYGNGAHSLTSLKEASTIAYISGYORS 89
DB 3 CPKGFMDRSGCYGFGELTRKAEAMCKVIHACGLASHLSPHEHAVALAFKFPQR 62
QY 90 QP---TWIGLHDPKROQMO---WIDGAMYLRSWSGKMGNGHCAEMSSNNPLTWS 142
DB 63 EEDWVWVIGLH-----HMNQARVWIDSGKKRYSAMDDDELPRGKYCTVLSGSGFME 116
QY 143 SNECKRQHFLCKY 156
DB 117 DNACSEBNPFCVKY 130

RESULT 10
LITL MOUSE
LITL MOUSE STANDARD; PRT; 165 AA.
AC P43137;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Lithostathine 1 precursor (pancreatic stone protein 1) (PSP)
DE (pancreatic thread protein 1) (PTP) (Islet of Langerhans regenerating
DE protein 1) (REG 1).
GN Name=Reg1;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=93340209; PubMed=8340418;
RA Ueno M., Yonekura H., Nakagawara K.-I., Watanabe T., Miyashita H.,
RA Morizumi S., Okamoto H., Itoh T., Teraoka H.;
RT "Structure, chromosomal localization, and expression of mouse reg
RT genes, reg I and reg II. A novel type of reg gene, reg II, exists in
RT the mouse genome."
RL J. Biol. Chem. 268:15974-15982(1993).
RN 12
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Struhsberg R.L., Feligold E.A., Grouse I.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Alteschul S.F., Zeeberg B., Buetow K.H., Moore T., Wang J., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Cantinici P., Prange C.,
RA Rahn S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.V., Malek J.A., Gunnaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton B., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Might act as an inhibitor of spontaneous calcium
CC carbonate precipitation.
CC -1- TISSUE SPECIFICITY: Expressed only in regenerating islets and

```

```

CC normal exocrine pancreas, but not in normal pancreatic islets.
CC Expressed strongly in pancreas, moderately in gallbladder, and
CC weakly in liver.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; D14010; BAA0311.1; -.
DR EMBL; BC028761; AA028761.1; -.
DR PIR; A47148; A47148.
DR HSSP; P05451; LITL.
DR MGD; MGI:97895; Reg1.
DR InterPro; IPR002353; Antifreeze1.
DR InterPro; IPR001304; Lectin C.
DR InterPro; IPR003990; Pancreatic_ac.
DR Pfam; PF00059; Lectin_C_1.
DR PRINTS; PR00356; ANTIFREEZE1.
DR PRINTS; PR01504; PNCBRATITSAP.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE LECTIN_2; 1.
KM Glycoprotein; Lectin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 165 Lithostathine 1.
FT DOMAIN 33 163 C-type lectin.
FT DISULFID 35 46 By similarity.
FT DISULFID 63 161 By similarity.
FT DISULFID 136 153 By similarity.
FT CARBOHYD 129 129 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 165 AA; 18518 MW; 2950174AF5D66BA CRC64;

Query Match
Beet Local Similarity 31.1%; Pred. No. 5.2e-17;
Matches 52; Conservative 35; Mismatches 51; Indels 29; Gaps 7;

QY 10 LILSLCLANTGYLADINRP-----SCAPGFYHNSCYGFRKLRMSDA 54
DB 8 ILLSCL-----IYLSPSQGEAEEDLPARISCEGSAVSYCYFIEDRLTWADA 59
QY 55 ELECGSYGNGAHSLTSLKEASTIAYI--SGYRQSPYTGILHDPKROQMOIWIDAM 112
DB 60 DLFCQNMNSG-YLVSVLSQAEGNFVASLTKESGTTDAN-VWTGLHDPKRRMRHWSGSL 117
QY 113 YLYRSM-SGKSMGNK-HCAEMSSNNPLTWSNECKRQHFLCKYR 157
DB 118 FLYSMATGSPNSNRKGVCLTSNTGYKKMKDNCDAQYFVCKFK 164

RESULT 11
LITL RAT
LITL RAT STANDARD; PRT; 165 AA.
AC P10758;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Lithostathine precursor (pancreatic stone protein) (PSP) (pancreatic
DE thread protein) (PTP) (Islet of Langerhans regenerating protein) (REG)
DE (Islet cells regeneration factor) (ICRF).
GN Name=Reg1; Synonyms=Reg;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=91093273; PubMed=1985964;
RA Rouquier S., Verdier J.M., Iovanna J., Dagorn J.-C., Glogori D.;

```


RT "Rat pancreatic stone protein messenger RNA. Abundant expression in
RT mature exocrine cells, regulation by food content, and sequence
RT identity with the endocrine reg transcript.";
RT J. Biol. Chem. 266:786-791 (1991).
RX SEQUENCE FROM N.A.
RX MEDLINE=88115343; PubMed=2963000;
RA Terazono K., Yamamoto H., Takaesawa S., Shiga K., Yonemura Y.,
RA Tochino Y., Okamoto H.;
RT "A novel gene activated in regenerating islets.";
RT J. Biol. Chem. 263:2111-2114 (1988).
RN (3)
RN SEQUENCE FROM N.A.
RX MEDLINE=93326645; PubMed=7916640; DOI=10.1016/0167-4781(93)90100-R;
RA Dubecq N.U., Fritgerio J.M., Dagorn J.-C., Iovanna J.L.;
RT "Rapid PCR cloning and sequence determination of the rat lithostathine
RT gene.";
RL Biochim. Biophys. Acta 1174:99-102 (1993).
RN (4)
RN SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RA Miyashita H., Suzuki Y., Watanabe T., Umno M., Morizumi S.,
RA Yonekura H., Okamoto H.;
RT "Structure and characterization of rat Reg I gene.";
RL Selfgaku 65:1082-1082 (1993).
RN (5)
RN SEQUENCE OF 22-69.
RC TISSUE=Pancreeas;
RX MEDLINE=90031455; PubMed=2680252;
RA Adrich Z., de Caro A.M., Guidoni A.A., Mouderra M.E., Rovey M.;
RT "Characterization in rat pancreatic juice of a protein homologous to
RT the human pancreatic stone protein.";
RL Comp. Biochem. Physiol. 93B:793-797 (1989).
CC -1- FUNCTION: Might act as an inhibitor of spontaneous calcium
CC carbonate precipitation.
CC TISSUE SPECIFICITY: Expressed only in regenerating islets, but not
CC in normal pancreatic islets, insulinomas or regenerating liver.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L07512; AAA1533.1; -
DR EMBL; M62930; AAA1974.1; -
DR EMBL; M18962; AAA2028.1; -
DR EMBL; D26164; BAA05149.1; -
DR PIR; A28351; A28351.
DR HSSP; P05451; 1LIT.
DR RGD; 3552; RegI.
DR InterPro: IPR001390; Lectin C.
DR InterPro: IPR003990; Pancreatic_ac.
DR Pfam; PF00059; Lectin_C; 1.
DR PRINTS; PR01504; PNCREATITAP.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_2; 1.
KW Direct protein sequencing; Glycoprotein; Lectin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 165
FT DOMAIN 33 163
FT DISULFID 35 146
FT DISULFID 63 161
FT DISULFID 136 153
FT CARBOHYD 129 129
SQ SEQUENCE 165 AA; 18672 MW; 9B61EB336B82CF8A CRC64;
Query Match 29.0%; Score 254.5; DB 1; Length 165;
Best Local Similarity 30.6%; Pred. No. 5.2e-17;

Matches 52; Conservative 38; Mismatches 51; Indels 29; Gaps 7;
QY 7 RLLILISCIATGVADITMRPS-----CAPWFTHKSNCGYFRKLAW 51
DB 5 KYFILLSTCL-----WVLSFSPQGEAREDDPSARITCEGSNNVSSYCYFMEHDLSW 56
QY 52 SAAELSCOSYNGAHASLISLKEASTIAEYI--SGYQSQPIWGLHPQKQOQWID 109
DB 57 AEADLFCCOMNSG-YIVSVLSQAEGNFSLIKESGTTAN-VWIGLHPKNNRRHWS 114
QY 110 GAMYLFRSW-SGKSMGK-NHCAEMSSNNPFTWSNNECKRQHFCKYR 157
DB 115 GSLFLYKSWDTGYPNNSNRGCVSVTSNSGCKKMDNSDQSLFVCKRK 164
RESULT 12
ID STR2_STRCA STANDARD; PRT; 142 AA.
AC P83515;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Struthioalcin-2 (SCA-2).
OS Struthio camelus (Ostrich).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Struthioniformes; Struthionidae;
OC Struthio.
OX NCBI_TaxID=8801;
RN (1)
RN SEQUENCE, SUBCELLULAR LOCATION, PHOSPHORYLATION SITES SER-62; SER-66
RP AND SER-68, AND MASS SPECTROMETRY.
RC TISSUE=Eggshell matrix;
RX PubMed=14726203; DOI=10.1016/j.bbapap.2003.09.006;
RA Mann K., Siedler F.;
RT "Ostrich (Struthio camelus) eggshell matrix contains two different C-
RT type lectin-like proteins. Isolation, amino acid sequence, and
RT posttranslational modifications.";
RL Biochim. Biophys. Acta 1696:41-50 (2004).
CC -1- SUBCELLULAR LOCATION: Eggshell matrix.
CC -1- MASS SPECTROMETRY: MW=16834.1; MW_ERR=2; METHOD=Electrospray;
CC RANGER=1-142; NOTE=Ref.1.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC InterPro: IPR001390; Lectin_C.
CC InterPro: IPR003990; Pancreatic_ac.
CC Pfam; PF00059; Lectin_C; 1.
CC PRINTS; PR01504; PNCREATITAP.
CC SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; PALSE_NEG.
DR PROSITE; PS00615; C_TYPE_LLECTIN_2; 1.
KW Direct protein sequencing; Lectin; Phosphorylation.
FT DOMAIN 13 139
FT DISULFID 6 17
FT DISULFID 34 138
FT DISULFID 113 130
FT MOD_RES 62 62
FT MOD_RES 66 66
FT MOD_RES 68 68
FT MOD_RES 68 68
SQ SEQUENCE 142 AA; 16601 MW; AF9950BF166B8FF9 CRC64;
Query Match 38.1%; Score 247; DB 1; Length 142;
Best Local Similarity 34.5%; Pred. No. 2.4e-16;
Matches 48; Conservative 17; Mismatches 66; Indels 8; Gaps 1;
QY 27 RPSCAPGWFHYKSNCGYFRKLAWNSDALEQOSYNGAHASLISLKEASTIAEYISGY 86
DB 3 RAGCAKGMWIPFDGRGCFPPQELSWRABGFCQRIQARTHLASISSEBHQIVSMNNS 62
QY 87 Q-----RSPVITIGLHDPKQKQOQWIDAMLYSKSKSGKNGKNGKLEMSSNNF 138
DB 63 QYSDSEEBAGEEVMWIGLHPIGRNWEWSDETGLDYSWYDVLRRRAVCLADDTTDF 122
139 LTWSSNECKRQHFCKYR 157
| : : : | : : : |

DB 123 ATWDELCSDRKPFICEYR 141

RESULT 13

PAP3_MOUSE STANDARD; PRT; 174 AA.

AC 009049;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Pancreatitis-associated protein 3 precursor (REG III-gamma).

GN Name=Pap3; Synonyms=Reg3g;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Liver, and Pancreas;

RX MEDLINE=97208868; PubMed=9055810; DOI=10.1016/S0378-1119(96)00589-6;

RA Narushima Y., Umno M., Nakagawa K.-I., Mori M., Miyashita H.,

RA Suzuki Y., Noguchi N., Takasawa S., Kumagai T., Yonekura H.,

RA Okamoto H.;

RT "Structure, chromosomal localization and expression of mouse genes encoding type III Reg, RegIII alpha, RegIII beta, RegIII gamma.";

RL Gene 185:159-168(1997).

CC -!- FUNCTION: Might be a stress protein involved in the control of bacterial proliferation.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- TISSUE SPECIFICITY: Constitutively expressed in the small intestine, moderately in colon and at an extremely low level in healthy pancreas.

CC -!- SIMILARITY: Contains 1 C-type lectin family domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL; D63361; BAA18930.1; -.

DR EMBL; D63362; BAA18931.1; -.

DR HSSP; P05451; 1LIT.

DR MGD; MGI:109406; Reg3g.

DR InterPro; IPR001304; Lectin C.

DR InterPro; IPR003990; Pancreatis_ac.

DR Pfam; PF00059; Lectin C; 1.

DR PRINTS; PR01504; PNCREATITSAP.

DR SMART; SM00034; CLECT; 1.

DR PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.

DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.

KW Acute phase; Inflammatory response; Lectin; Multigene family; signal.

FT SIGNAL 1 26 Potential.

FT CHAIN 27 174 Pancreatitis-associated protein 3.

FT DOMAIN 38 174 C-type lectin.

FT DISULFID 40 51 By similarity.

FT DISULFID 68 170 By similarity.

FT DISULFID 145 162 By similarity.

FT SEQUENCE 174 AA; 19307 MW; 5575B956A4D8CEP CRC64;

Query Match 28.1%; Score 246.5; DB 1; Length 174;

Best Local Similarity 35.2%; Pred. No. 3.4e-16;

Matches 57; Conservative 24; Mismatches 66; Indels 15; Gaps 6;

QY 10 LLLSCL-AKTGVLDII-----MRPSCAPGMFYHKSNCYGYFRKILRMWSDALELCQSYG 62

DB 13 MLTSLTMLSLQYGVAKKADAPSRSRSCPKSGRAAGSYALFVSXKMYDMDMACQKRP 72

QY 63 NGAHLSLTLKASTIAEYI-SGYORSQPIWIGLHP-----QKROQWQIDGAMLYR 116

DB 73 SG-HLVSVYLSGAELASFLSSMKISGNSGQYVWIGLHDPITLGYEPNRGGWEMSNADVNYI 131

QY 117 SW-SGKSGKNGKHCAMSSNNNFLTWSNECKRQHFLCKYR 157

DB 132 NWETNPSSSGNHCGCTLSRASGLFKMKNYCNLELPYCKRK 173

RESULT 14

O6TR86 PRELIMINARY; PRT; 164 AA.

AC O6TR86;

DT 05-JUL-2004 (TREMBlrel. 27, Created)

DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)

DE Btcul.

OS Bothrops jararacusu (Jararacusu).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Lepidodactylus; Squamata; Scleroglossa; Serpentes; Colubroidae;

OC Viperidae; Crotalinae; Bothrops.

OX NCBI_TaxID=8726;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Venom;

RX PubMed=15135412; DOI=10.1016/j.pep.2004.02.012;

RA Kaasab B.H., de Carvalho D.D., Oliveira M.A., Baptista G.R.,

RA Pereira G.A., Novello J.C.;

RT "Cloning, expression, and structural analysis of recombinant Btcul, a c-type lectin from the Bothrops jararacusu snake venom.";

RL Protein Expr. Purif. 35:344-352(2004).

DR EMBL; AY388642; AAQ2957.1; -.

DR GO; GO:0005529; F:sugar binding; IEA.

DR InterPro; IPR001304; Lectin C.

DR InterPro; IPR003990; Pancreatis_ac.

DR Pfam; PF00059; Lectin C; 1.

DR PRINTS; PR01504; PNCREATITSAP.

DR SMART; SM00034; CLECT; 1.

DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.

DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.

FT CHAIN 24 157 Btcul.

FT SEQUENCE 164 AA; 19070 MW; F88101D4338B94DC CRC64;

Query Match 27.9%; Score 245; DB 2; Length 164;

Best Local Similarity 36.9%; Pred. No. 4.5e-16;

Matches 48; Conservative 16; Mismatches 62; Indels 4; Gaps 2;

QY 29 SCAPGMFYHKSNCYGYFRKILRMWSDALELCQSYNGAHLSLTKASTIAEYISGYOR 88

DB 25 NCPQDMLPMNGLCYKIFNELKAMDAEMFCKRYKPGCHLASIHLYGESPREIAEYISDYHK 84

QY 89 SQ-PIWIGLHPKROQWQIDGAMLYRSWSGKS---MGNGKHCAMSSNNNFLTWSNN 144

DB 85 GQSEWVIGLQCKKDFSWEMWDRSCDTYLSMDKNQPDHYQKRCVELSVSTGYRLMNDQ 144

QY 145 ECKRQHFLC 154

DB 145 VCESKNAFLC 154

RESULT 15

O6OX33 PRELIMINARY; PRT; 158 AA.

AC O6OX33;

DT 05-JUL-2004 (TREMBlrel. 27, Created)

DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)

DE C-type lectin.

OS Bothrops insularis (Island jararaca) (Queinada jararaca).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Lepidodactylus; Squamata; Scleroglossa; Serpentes; Colubroidae;

OC Viperidae; Crotalinae; Bothrops.

OX NCBI_TaxID=8723;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Venom gland;

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2005, 23:32:47 ; Search time 39 Seconds

(without alignments)
389,801 Million cell updates/sec

Title: US-09-525-041-2

Perfect score: 878

Sequence: 1 MARSRLLLLSCLAKTGV.....LTWSSNECKRQHFICKTRP 158

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:*

1: D471:*

2: P12:*

3: P13:*

4: P14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	254.5	29.0	165	A47148	reg I, regenerating
2	254.5	29.0	165	A28351	pancreatic stone p
3	241.5	27.5	142	S78596	ovocleidin - chick
4	240	27.3	166	RGHUIB	regenerating islet
5	236.5	26.9	173	B47148	reg II, regenerati
6	236.5	26.9	174	I83377	regenerating prote
7	233	26.5	135	A38609	lectin, galactose-
8	231	26.3	174	S54979	pancreatitis-asso
9	221	26.3	175	A37194	pancreatic thread
10	228.5	26.0	174	A48689	pancreatitis-asso
11	225.5	25.7	166	RGHUIA	regenerating islet
12	224	25.5	166	A45751	pancreatic stone p
13	222.5	25.3	172	S32489	lectin - Iberian r
14	221	25.2	175	A49616	pancreatitis-asso
15	206.5	23.5	152	UC7134	agglutinin alpha
16	203.5	23.2	131	UC5058	bitiscetin alpha c
17	203	23.1	175	A41719	pancreatic stone p
18	202.5	23.1	123	UC2415	echinectin beta cha
19	202.5	23.1	152	UC4690	coagulation factor
20	200.5	22.8	125	UC5059	bitiscetin beta cha
21	196.5	22.4	146	UC7105	aggregin beta cha
22	196.5	22.4	146	UC4691	coagulation factor
23	194	22.1	175	S29822	pancreatitis-asso
24	193	22.0	330	T46256	brevican - human
25	193	22.0	312	A54423	brevican precursor
26	191	21.8	883	S57653	brevican precursor
27	189	21.5	133	A47267	botrocetin alpha c
28	185.5	21.1	125	B47267	botrocetin beta cha
29	185.5	21.1	129	UC4329	coagulation factor

30	185	21.1	883	2	S49126	brevican precursor
31	185	21.1	1257	2	S28764	neurocan precursor
32	184.5	21.0	146	2	UC7135	agglutinin beta
33	182.5	20.8	301	2	S13165	asialoglycoprotein
34	182	20.7	144	2	PC7027	aggregin alpha cha
35	181	20.6	291	1	LNHU1	hepatic lectin H1
36	181	20.6	3562	2	A47171	chondroitin sulfat
37	179	20.4	1268	2	S52781	neurocan - mouse
38	179	20.4	2397	1	A55535	versican precursor
39	179	20.4	2409	1	A60979	versican precursor
40	178.5	20.3	311	1	LNHU2A	asialoglycoprotein
41	177.5	20.2	162	1	LNRC1	lectin BRA3-1 prec
42	177.5	20.2	162	1	LNRC3	lectin BRA3-2 prec
43	177	20.2	1643	2	T14274	versican precursor
44	177	20.2	3381	2	T42389	versican precursor
45	172.5	19.6	123	2	B42972	coagulation factor

ALIGNMENTS

RESULT 1
A47148
reg I, regenerating islet cells - mouse
C.Species: Mus musculus (house mouse)
C.Date: 03-May-1994 #sequence revision 03-May-1994 #text_change 09-Jul-2004
C.Accession: A47148
R.Umno, M.; Yonekura, H.; Nakagawa, K.; Watanabe, T.; Miyashita, H.; Morizumi, S.; Ok
J. Biol. Chem. 268, 15974-15982, 1993
A.Title: Structure, chromosomal localization, and expression of mouse reg genes, reg I a
A.Reference number: A47148; MUID:93340209; PMID:8340418
A.Accession: A47148
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-165 <UNN>
A.Cross-references: UNIPROT:P43137; GB:D14010; NID:G391771; PIDN:BA03111.1; PID:G391772
A.Intons: 21/1, 60/3, 106/3, 144/1
C.Superfamily: tetralectin; C-type lectin homology
P.35-161/Domain: C-type lectin homology <LCH>
P.35-46/63-161,136-153/Dileufide bonds: #status predicted

Query Match 29.0%; Score 254.5; DB 2; Length 165;
Best Local Similarity 31.1%; Pred. No. 9,2e-18;
Matches 52; Conservative 35; Mismatches 51; Indels 29; Gaps 7;

QY	10	LLLSCLAKTGVLDIIRP-----SCAPGWFYHKSCYGFPRKLRNWSDA	54
DB	8	ILLSCL-----IVLSPQGEAEEDLPKARICCPGSSNAYSCYFTEDRLTWMDA	59
QY	55	ELTEGSGYNGAHLASILKEASTAEYI--SGYRSQPIWIGLHPQKQOWIDGAM	112
DB	60	DLFCQNMMSG-YLVSVLSQAEQNFVASILKESGTTDAN-VMTGLDHPKRNRRHWSGSL	117
QY	113	VYRSG-SGKSMGK-NHCAEMSSNNPLTWSNCKRQHFICKYR	157
DB	118	FLYKSWATGSPNNSNRGYCVSLTSMYTKKWDNDCAQYSPVCKP	164

RESULT 2
A28351
pancreatic stone protein precursor - rat
N.Alternate names: lithoecathine
C.Species: Rattus norvegicus (Norway rat)
C.Date: 31-Mar-1990 #sequence revision 31-Mar-1990 #text_change 09-Jul-2004
C.Accession: A28351; A39081; P10147; S34618
J.Rterazono, K.; Yamamoto, H.; Takasawa, S.; Shiga, K.; Yonemura, Y.; Tochino, Y.; Okamoto
J. Biol. Chem. 263, 2111-2114, 1988
A.Title: A novel gene activated in regenerating islets.
A.Reference number: A92704; MUID:88115343; PMID:2963000
A.Accession: A28351
A.Molecule type: mRNA
A.Residues: 1-165 <TER>

A:Cross-references: UNIPROT:P10758; GB:M18963; NID:g206604; PIDN:AAA42028.1; PID:g206603
R.Rouquier, S.; Vergier, J.M.; Iovanna, J.; Dagorn, J.C.; Glogri, D.
J. Biol. Chem. 266, 786-791, 1991

A>Title: Rat pancreatic stone protein messenger RNA. Abundant expression in mature exocrine pancreas.

A:Reference number: A39081; MUID:91093273; PMID:1985964

A:Accession: A39081

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-165 <ROU>

A:Cross-references: GB:M62930; GB:J05722; NID:g206462; PIDN:AAA41974.1; PID:g206463
R.Adrich, Z.; De Caro, A.M.; Guidoni, A.A.; Woudstra, M.E.; Rovey, M.
Comp. Biochem. Physiol. B 93, 793-797, 1989

A>Title: Characterization in rat pancreatic juice of a protein homologous to the human P10147

A:Reference number: P10147; MUID:90031455; PMID:2680252

A:Accession: P10147

A:Molecule type: protein

A:Residues: 22-69 <ADR>

A:Experimental source: pancreas

R.Dusetet, N.J.; Frigerio, J.M.; Dagorn, J.C.; Iovanna, J.L.
Biochim. Biophys. Acta 1174, 99-102, 1993

A>Title: Rapid PCR cloning and sequence determination of the rat lithostathine gene.

A:Reference number: S34618; MUID:93326645; PMID:7916640

A:Accession: S34618

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-165 <DUS>

A:Cross-references: EMBL:L07512; NID:g393208; PIDN:AAA1533.1; PID:g393209
C.Comment: This protein is found in pancreatic calculi of mammals. A peptide bond between
residue 1 into an insoluble protein at a neutral pH of 5.5 to 7.5.

C:Genetics:

A:Introns: 21/1; 60/3; 106/3; 144/1

C:Superfamily: tetraneurin; C-type lectin homology

C:Keywords: pyroglutamic acid

F1-21/Domin: signal sequence #status predicted <SIG>

F122-165/Product: pancreatic stone protein #status predicted <MAT>

F122-165/Domin: C-type lectin homology <LCH>

F122/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental

Query Match	27.5%	Score 24.5	DB 2	Length 142
Beet Local Similarity	33.8%	Pred. No. 1.5e-16		
Matches	46	Conservative 23	Mismatches 56	Indels 11
			Gaps	3
QY	30	CAPGFWFKNSCYGFKRLRNWSDALECQSYGNGAHLSILTKASTIAEYI-----	83	
DB	5	CGPGVPPVPGGLGFGFSRELSSWRLESFPRKRGPGSHLAAYNSAALLRLRLAEILNNSRG	64	
QY	84	--SGYRSQPIWIGLHDPKRRQOWIDGAMYLNS--SGKSMGNGKICAEKSSNNFL	139	
DB	65	DGSGGAGDGRVWIGHRPAGSRSMWSDGTAPRFASWHTAYARRGR--CAALRDEEAF	123	
QY	140	TWSSNECKRQHFLCK	155	
DB	124	SWAAPRTERRNAFVK	139	

RESULT 4
RGHUB

regenerating islet lectin 1-beta precursor - human
N;Alternate names: reg-related protein; reg1-beta protein
N;Contains: pancreatic stone protein (PSP)

C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence revision 03-Aug-1995 #text change 09-Jul-2004

C/Accession: S34591; S42729; A44712
R/Bertoli, C.; Ghaleb, B.; Giorgi, D.; Sansonetti, A.; Dagorn, J.C.; Berge-Lefranc, J.L.

FEBS Lett. 327, 289-293, 1993

A;Reference number: S34591; MUID:93351647; PMID:8348956
A;Accession: S34591

A;Molecule type: DNA
A;Residues: 1-166 <BAR>

A/Note: this gene appears to be expressed in pancreas and liver

R; Morizumi, S.; Watanabe, T.; Umno, M.; Nakagawara, K.; Suzuki, Y.; Miyashita, H.; Yonekura, M.; Yonekura, Y. *Biochim. Biophys. Acta* 1217, 199-202, 1994

A/Title: Isolation, structural determination and expression of a novel reg gene, human
A/Reference number: S42729; MUID:94153997; PMID:8110835

A/Accession: S42729
A/Molecule type: mRNA

A;Residues: 1-166 <MOR>
A;Cross-references: GB:D16816; NID:g474305; PIDN:BAA04091.1; PID:g474306

A/Accession: A44712
A/Molecule type: DNA

A/Residues: 1-166 <MO2>
A/Cross-references: GB:DJ17291; NID:g474307; PIDN:BAA04124.1; PID:g474308

C/Comment: This protein is found in pancreatic calculi of mammals. The tryptic-like cleavage site is at the C-terminus.

A;Gene: GDB:REG1B; REG1
A;Cross-references: GDB:342079

A;Map position: 2p12-2p12
A;Introns: 22/1; 61/3; 107/3; 145/1

C: Superfamily: tetranectin; C-type lectin homology
C/keywords: glycoprotein; lectin; pancreas; pyroglutamic acid

F;1-22/Domain: signal sequence #status predicted <516>
F;23-166/Product: regenerating islet lectin Ibeta #status predicted <NAT>

F;34-166/Product: pancreatic stone protein #status predicted <MALZ>
F;36-162/Domain: C-type lectin homology <LGH>

F/23/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F/27/Binding site: carbohydrate (Thr) (covalent) #status predicted

```
F;33-34/Cleavage site: Arg-11e (trypsin) #status predicted
F;36-47,64-162,137-154/Disulfide bonds: #status predicted
```

Query Match	Score	DB	Length
27.3%	240	1	166

Best Local Similarity 34.3%; Pred. NO. 2.5e-16;
Matches 46; Conservative 27; Mismatches 57; Indels 4; Gaps 4;

27 RPSCAPGWFYHKSNVCYGYFRKLRNMSDAIECQSYNGAHLASTLSLKEASTIAEYISGY 86

Db 33 RISCPEGTNAYRSYCYFFNEDEPETWDADLYCQNNMSG-NLVSVLQAEGAFVSLIKES 91

QY 87 QRSQ- IWIGHDPKROQWOWIDGANYLYRSW-SGKSGGNK-HCAEMSSNNPFLTWS 143

Db 92 STDDSNVWIGLHDPKRRRHHMSSGSLVYKSWDTGSSPSSANNGYCASLTSCSGFKMKMD 151

Qy 144 NECNKRQHFLCKR 157
152 ESCEKFFSVCKFK 165

RESULT 5
B47148
reg II, regenerating islet cells - mouse
C.Species: Mus musculus (house mouse)
C.Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C.Accession: B47148
R.Umno, M.; Yonekura, H.; Nakagawara, K.; Watanabe, T.; Miyashita, H.; Morizumi, S.; Ok
J. Biol. Chem. 268, 15974-15982, 1993
A.Title: Structure, chromosomal localization, and expression of mouse reg genes, reg I &
A.Reference number: A47148; MUID:93340209; PMID:8340418
A.Accession: B47148
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-173 <UNN>
A.Cross-references: UNIPROT:Q08731; GB:D14011; NID:9391773; PIDN:BAA03112.1; PID:9391774
C.Genetic:
A.Introns: 22/1; 68/3; 114/3; 152/1
C.Superfamily: tetranectin; C-type lectin homology
F.43.169/Domain: C-type lectin homology <LCH>
F.43-54,71-169,144-161/Disulfide bonds: #status predicted

Query Match 26.9%; Score 236.5; DB 2; Length 173;
Best Local Similarity 28.7%; Pred. No. 5.8e-16;
Matches 50; Conservative 33; Mismatches 72; Indels 19; Gaps 5;

Qy 1 MASRNRLLLSLCAKTGVLDTI-----MRPSCAPGFYHKSNCTCYFRK 47
119 HNSGSLFLFKMAGAPSTANNGYCVSLSTNAYKMDENCEAGYFVCKFR 172

Db 1 MANNVYLLFLCLMFLSYSGQVAREDEPLAEKDLPSAKINCPGANNVGSYCYLLID 60
48 LNWSDAELECSYGNAGHASTLSLKEASTIAEYI--SGYRSGQPIWGLHDPQKQQM 105
61 RLTWGADLPQQM--NAGHIVSLTSLQESNFAVSLVKESTGTSN--VWGLHDPKSNRRM 118

Qy 106 QWIDGAMLYRSWS--GKSMGNKHCAMSSNNNFLTWSNECNKRQHFLCKR 157
119 HNSGSLFLFKMAGAPSTANNGYCVSLSTNAYKMDENCEAGYFVCKFR 172

Db 119 HNSGSLFLFKMAGAPSTANNGYCVSLSTNAYKMDENCEAGYFVCKFR 172

RESULT 6
183377
regenerating protein III (reg III) - rat
C.Species: Rattus norvegicus (Norway rat)
C.Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
C.Accession: I60296; 183377
R.Suzuki, Y.; Yonekura, H.; Watanabe, T.; Umno, M.; Morizumi, S.; Miyashita, H.; Okamoto
Gene 144, 315-316, 1994
A.Title: Structure and expression of a novel rat RegIII gene.
A.Reference number: I60296; MUID:94314238; PMID:8039722
A.Accession: I60296
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-174 <RES>
A.Cross-references: UNIPROT:P35231; GB:D23676; NID:9471157; PIDN:BAA04904.1; PID:9471158
A.Accession: I83377
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-174 <RES>
A.Cross-references: GB:D26078; NID:9471159; PIDN:BAA05071.1; PID:9471160
A.Genetic:
A.Introns: 25/1; 64/3; 110/3; 153/1
C.Superfamily: tetranectin; C-type lectin homology
F.39-170/Domain: C-type lectin homology <LCH>

Query Match 26.9%; Score 236.5; DB 2; Length 174;
Best Local Similarity 35.3%; Pred. No. 5.8e-16;

Matches 49; Conservative 25; Mismatches 56; Indels 9; Gaps 4;

Qy 27 RPSCAPGFYHKSNCTCYFRKLRWSDAELECSYGNAGHASTLSLKEASTIAEYISG 85
36 RSCPMGSKAYNSYCYTLVTLTKSWFOADLCQKRPSC-HLVSLISGEASFVSSLVYGR 94

Qy 86 YRSQPIWGLHDPQKQQ-----QWIDGAMLYRSWSG--KSMGNKHCAMSSNNNF 138
95 VANNODIWLHDPFMGQQPNCGGHMSNDVLYNTLMDGPSSIVYNRNGCSLTATSEF 154

Db 139 LTWSSNECNKRQHFLCKR 157
155 LKMGDHCDVLELPFCYCKFR 173

RESULT 7
A38609
lectin, galactose-specific - western diamondback rattlesnake
C.Species: Crotalus atrox (western diamondback rattlesnake)
C.Date: 23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change 09-Jul-2004
C.Accession: A38609
R.Hirabayashi, J.; Kusunoki, T.; Kasai, K.
J. Biol. Chem. 266, 2320-2326, 1991
A.Title: Complete primary structure of a galactose-specific lectin from the venom of the
A.Reference number: A38609; MUID:91115849; PMID:1989986
A.Accession: A38609
A.Status: preliminary
A.Molecule type: protein
A.Residues: 1-135 <HIR>
A.Cross-references: UNIPROT:P21963
C.Superfamily: tetranectin; C-type lectin homology
F.3-131/Domain: C-type lectin homology <LCH>
F.3-14,31-131,106-123/Disulfide bonds: #status predicted

Query Match 26.5%; Score 233; DB 2; Length 135;
Best Local Similarity 34.4%; Pred. No. 9.8e-16;
Matches 45; Conservative 18; Mismatches 64; Indels 4; Gaps 2;

Qy 29 SCAPGFYHKSNCTCYFRKLRWSDAELECSYGNAGHASTLSLKEASTIAEYISG 88
2 NCPDLWLPWNGLCYIFNQLKTWEDAEWFCRYKPKGCHLASFRYGESLAEYISDYHK 61

Db 89 SQP-IWGLHDPQKQQWQWIDGAMLYRSWSGKS--MGKHCAMSSNNNFLTWSN 144
62 GQENWIGLDRKKQPSWETDRSCTDYLTMDKQDPHYKQKPFVELVSLTGYRLMDQ 121

Qy 145 ECNKRQHFLCK 155
122 VCESDAFLCQ 132

Db 122 VCESDAFLCQ 132

RESULT 8
S54979
pancreatitis-associated protein PAP-3 - rat
C.Species: Rattus norvegicus (Norway rat)
C.Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
C.Accession: S54979; S43458
R.Duseti, N.J.; Frigerio, J.M.; Szpirer, C.; Dagorn, J.C.; Iovanna, J.L.
Biochem. J. 307, 9-16, 1995
A.Title: Cloning, expression and chromosomal localization of the rat pancreatitis-associa
A.Reference number: S54979; MUID:95234061; PMID:7717998
A.Accession: S54979
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-174 <DUS>
A.Cross-references: UNIPROT:P42854; EMBL:U09193; NID:9483931; PIDN:AAA79231.1; PID:94839
R.Frigerio, J.M.; Duseti, N.J.; Garrido, P.; Dagorn, J.C.; Iovanna, J.L.
Biochim. Biophys. Acta 1216, 329-331, 1993
A.Title: The pancreatitis associated protein III (PAP III), a new member of the PAP gene
A.Reference number: S43458; MUID:94060113; PMID:8241280
A.Accession: S43458
A.Status: preliminary
A.Molecule type: mRNA

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2005, 23:21:46 ; Search time 49 Seconds
(without alignments)
1053.599 Million cell updates/sec

Title: US-09-525-041-2

Perfect score: 878
Sequence: 1 MASRMRLLLLSLCAKTGV.....LTWSNECKRQHPICKRP 158

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
17:	/cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/1/pubppa/US11_NEW_PUB.pep.*
19:	/cgn2_6/ptodata/1/pubppa/US11_NEW_PUB.pep.*
20:	/cgn2_6/ptodata/1/pubppa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	878	100.0	158	US-09-922-217-1070	Sequence 1070, Ap
2	878	100.0	158	US-09-922-217-1077	Sequence 1077, Ap
3	878	100.0	158	US-09-922-217-1078	Sequence 1078, Ap
4	878	100.0	158	US-09-922-217-1079	Sequence 1079, Ap
5	878	100.0	158	US-09-922-217-1080	Sequence 1080, Ap
6	878	100.0	158	US-09-922-217-1080	Sequence 1080, Ap
7	878	100.0	158	US-09-922-217-1080	Sequence 1080, Ap
8	878	100.0	158	US-09-922-217-1080	Sequence 1080, Ap
9	878	100.0	158	US-09-922-217-1080	Sequence 1080, Ap
10	878	100.0	158	US-09-922-217-1080	Sequence 1080, Ap
11	878	100.0	158	US-09-922-217-1080	Sequence 1080, Ap
12	878	100.0	158	US-09-922-217-1080	Sequence 1080, Ap
13	878	100.0	158	US-09-922-217-1080	Sequence 1080, Ap

14	878	100.0	158	US-10-025-380-1077	Sequence 1077, Ap
15	878	100.0	158	US-10-025-380-1078	Sequence 1078, Ap
16	878	100.0	158	US-10-025-380-1079	Sequence 1079, Ap
17	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
18	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
19	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
20	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
21	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
22	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
23	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
24	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
25	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
26	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
27	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
28	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
29	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
30	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
31	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
32	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
33	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
34	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
35	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
36	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
37	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
38	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
39	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
40	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
41	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
42	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
43	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
44	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap
45	878	100.0	158	US-10-025-380-1080	Sequence 1080, Ap

ALIGNMENTS

RESULT 1
US-09-922-217-1070
Sequence 1070, Application US/09922217
Patent No. US/0020764141
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secretist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeline Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongrong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole Lynn
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C13
CURRENT APPLICATION NUMBER: US/09/922,217
NUMBER OF SEQ ID NOS: 1124
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1070
LENGTH: 158
TYPE: PRT
ORGANISM: Homo sapiens
US-09-922-217-1070

Query Match 100.0%; Score 878; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MASRMRLLLLSLCAKTGVLDIMRPSCAPGWFYHNSNCYGRKLRMSDALEECOS 60
|||||

Db 1 MASRSMRLLLLSCLACTGVLDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEBQCS 60
QY 61 YGNGAHLASITSLKEASTIAEYISGYORSQPIWIGLHDPQRQOWIDGAMTYLYRSWSG 120
Db 61 YGNGAHLASITSLKEASTIAEYISGYORSQPIWIGLHDPQRQOWIDGAMTYLYRSWSG 120
QY 121 KSMGNKHCAMSSNNNFLTWSNNECNKRQHFLECKYRP 158
Db 121 KSMGNKHCAMSSNNNFLTWSNNECNKRQHFLECKYRP 158

RESULT 2

US-09-922-217-1077
; Sequence 1077, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secretist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1077
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1077

Query Match 100.0%; Score 878; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRSMRLLLLSCLACTGVLDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEBQCS 60
Db 1 MASRSMRLLLLSCLACTGVLDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEBQCS 60
QY 61 YGNGAHLASITSLKEASTIAEYISGYORSQPIWIGLHDPQRQOWIDGAMTYLYRSWSG 120
Db 61 YGNGAHLASITSLKEASTIAEYISGYORSQPIWIGLHDPQRQOWIDGAMTYLYRSWSG 120
QY 121 KSMGNKHCAMSSNNNFLTWSNNECNKRQHFLECKYRP 158
Db 121 KSMGNKHCAMSSNNNFLTWSNNECNKRQHFLECKYRP 158

RESULT 3

US-09-922-217-1078
; Sequence 1078, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secretist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.

; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1078
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1078

Query Match 100.0%; Score 878; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRSMRLLLLSCLACTGVLDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEBQCS 60
Db 1 MASRSMRLLLLSCLACTGVLDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEBQCS 60
QY 61 YGNGAHLASITSLKEASTIAEYISGYORSQPIWIGLHDPQRQOWIDGAMTYLYRSWSG 120
Db 61 YGNGAHLASITSLKEASTIAEYISGYORSQPIWIGLHDPQRQOWIDGAMTYLYRSWSG 120
QY 121 KSMGNKHCAMSSNNNFLTWSNNECNKRQHFLECKYRP 158
Db 121 KSMGNKHCAMSSNNNFLTWSNNECNKRQHFLECKYRP 158

RESULT 4

US-09-922-217-1079
; Sequence 1079, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secretist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1079
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1079

Query Match 100.0%; Score 878; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRSMRLLLLSCLACTGVLDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEBQCS 60
Db 1 MASRSMRLLLLSCLACTGVLDIIMRPSCAPGWFYHKSNCYGFRKLRNMSDALEBQCS 60
QY 61 YGNGAHLASITSLKEASTIAEYISGYORSQPIWIGLHDPQRQOWIDGAMTYLYRSWSG 120
Db 61 YGNGAHLASITSLKEASTIAEYISGYORSQPIWIGLHDPQRQOWIDGAMTYLYRSWSG 120

Qy	Db
121 KSMGSKGKCAEMSSNNFLTWSSNECNKRQHFLLCKRRP 158	121 KSMGSKGKCAEMSSNNFLTWSSNECNKRQHFLLCKRRP 158

RESULT 5
US-09-922-217-1080

```

: Sequence 1080, Application US/09922217
: Patent No. US20070076414A1
:
: GENERAL INFORMATION:
: APPLICANT: Xu, Jianshun
: APPLICANT: Lodes, Michael J.
: APPLICANT: Secretist, Heather
: APPLICANT: Benson, Darin R.
: APPLICANT: Meagher, Madeline Joy
: APPLICANT: Stolk, John A.
: APPLICANT: Wang, Tongtong
: APPLICANT: Jlang, Yujung
: APPLICANT: Smith, Carole Lynn
: APPLICANT: King, Gordon E.
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jonathan D.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
: TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
: FILE REFERENCE: 210121.471C13
: CURRENT APPLICATION NUMBER: US/09/922,217
: CURRENT FILING DATE: 2001-08-03
: NUMBER OF SEQ ID NOS: 1124
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1080
: LENGTH: 158
: TYPE: PRT
: ORGANISM: Homo sapiens
:
: US-09-922-217-1080

```

Query Match	100.0%	Score 878;	DB 9;	Length 158;
Best Local Similarly	100.0%	Pred. No. 3.8e-82;		
Matches 158;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	MARSRRRLILLISCLAKTGLGDIIMRPFCAAGWFFHNSNCYGYPKRLRNMSDALEECQS	60
Db	1	MARSRRRLILLISCLAKTGLGDIIMRPFCAAGWFFHNSNCYGYPRKLRNMSDALEECQS	60
QY	61	YNGGAHLASTLSLKEASTIAEYISGVGRORPFIWGHPBQRQOMOWIDGMYLYRSMSG	120
Db	61	YNGGAHLASTLSLKEASTIAEYISGVGRORPIFIWGHPDQRQOMOWIDGMYLYRSMSG	120
QY	121	KSMGNRHCEMSSNNNPLTWSSNECNKRQHFLCYRP	158
Db	121	KSMGNRGHCLEMSSNNNPLTWSSNECNKRQHFLCYRP	158

RESULT 6
US-09-833-263-1070

```

: Sequence 1070, Application US/09833263
: Patent No. US2002010547A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jonathan D.
: APPLICANT: Stolk, John A.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
: FILE REFERENCE: 210121..471012
: CURRENT APPLICATION NUMBER: US/09/833,263
: CURRENT FILING DATE: 2001-04-10
: NUMBER OF SEQ. ID NOS.: 1093
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ. ID NO. 1070
: LENGTH: 158
: TYPE: PR1
: ORGANISM: Homo sapiens

```

US-09-833-263-1070

Query Match	100.0%	Score 878;	DB 9;	Length 158;
Best Local Similarity	100.0%	Pred. No. 3.8e-82;		
Matches 158; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	MSRSRRLLLLSCLAKTGVLDIMRSCAPGPFYKSNICYGFRLLRMSPDALEECQS	60
Db	1	MSRSRRLLLLSCLAKTGVLDIMRSCAPGPFYKSNICYGFRLLRMSPDALEECQS	60
QY	61	YNGGHAHSLILKEASTIAYISGYQSRQPIWIGLHDPQKROQMWDGAMLYLSWSG	120
Db	61	YNGGHAHSLILKEASTIAYISGYQSRQPIWIGLHDPQKROQMWDGAMLYLSWSG	120
QY	121	KSWGKHCACEMSSNNNFLTWSNCECNTRQHFLCKYRP	158
Db	121	KSWGKHCACEMSSNNNFLTWSNCECNTRQHFLCKYRP	158

RESULT 7
US-09-833-263-1077

```

Sequence 1077, Application US/09833263
Patent No. US20020110547A1
GENERAL INFORMATION:
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Stolk, John A.
APPLICANT: Meagher, Madeleine J.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.477C12
CURRENT APPLICATION NUMBER: US/09/833,263
CURRENT FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 1093
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1077
LENGTH: 158
TYPE: PR1
ORGANISM: Homo sapiens
US-09-833-263-1077

```

Query Match	100.0%;	Score 878;	DB 9;	Length 158;
Best Local Similarity	100.0%;	Pred. No. 3.8e-82;		
Matches 158;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

RESULT 8

```

1 US-09-833-263-1078
2 Sequence 1078, Application US/09833263
3 Patent No. US20020110547A1
4 GENERAL INFORMATION:
5 APPLICANT: Wang, Aijun
6 APPLICANT: Clapper, Jonathan D.
7 APPLICANT: Stolk, John A.
8 APPLICANT: Meagher, Madeleine J.
9 TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
10 TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
11 FILE REFERENCE: 210121.471C12
12 CURRENT APPLICATION NUMBER: US/09/833,263
13 CURRENT FILING DATE: 2001-04-10
14 NUMBER OF SEQ ID NOS: 1093

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1078
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-263-1078

Query Match 100.0%; Score 878; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MASSRMRLLLLSCLAKTGVLDIIMRPSCAPGWFYHKSNCYGYFRKLKRWSDALEECQS 60
Db 1 MASSRMRLLLLSCLAKTGVLDIIMRPSCAPGWFYHKSNCYGYFRKLKRWSDALEECQS 60
Qy 61 YNGNAHLASIIISLKEASTIAEYISGYORSOPITWIGLHDPQKRQOWIDGAMLYLRWSWG 120
Db 61 YNGNAHLASIIISLKEASTIAEYISGYORSOPITWIGLHDPQKRQOWIDGAMLYLRWSWG 120
Qy 121 KSMGNKHCACEMSSNNNFLLTWSSNECNKRQHFLLCKYRP 158
Db 121 KSMGNKHCACEMSSNNNFLLTWSSNECNKRQHFLLCKYRP 158

RESULT 9
US-09-833-263-1079
; Sequence 1079, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833,263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1079
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-263-1079

Query Match 100.0%; Score 878; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MASSRMRLLLLSCLAKTGVLDIIMRPSCAPGWFYHKSNCYGYFRKLKRWSDALEECQS 60
Db 1 MASSRMRLLLLSCLAKTGVLDIIMRPSCAPGWFYHKSNCYGYFRKLKRWSDALEECQS 60
Qy 61 YNGNAHLASIIISLKEASTIAEYISGYORSOPITWIGLHDPQKRQOWIDGAMLYLRWSWG 120
Db 61 YNGNAHLASIIISLKEASTIAEYISGYORSOPITWIGLHDPQKRQOWIDGAMLYLRWSWG 120
Qy 121 KSMGNKHCACEMSSNNNFLLTWSSNECNKRQHFLLCKYRP 158
Db 121 KSMGNKHCACEMSSNNNFLLTWSSNECNKRQHFLLCKYRP 158

RESULT 10
US-09-833-263-1080
; Sequence 1080, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833,263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1080
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-263-1080

Query Match 100.0%; Score 878; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MASSRMRLLLLSCLAKTGVLDIIMRPSCAPGWFYHKSNCYGYFRKLKRWSDALEECQS 60
Db 1 MASSRMRLLLLSCLAKTGVLDIIMRPSCAPGWFYHKSNCYGYFRKLKRWSDALEECQS 60
Qy 61 YNGNAHLASIIISLKEASTIAEYISGYORSOPITWIGLHDPQKRQOWIDGAMLYLRWSWG 120
Db 61 YNGNAHLASIIISLKEASTIAEYISGYORSOPITWIGLHDPQKRQOWIDGAMLYLRWSWG 120
Qy 121 KSMGNKHCACEMSSNNNFLLTWSSNECNKRQHFLLCKYRP 158
Db 121 KSMGNKHCACEMSSNNNFLLTWSSNECNKRQHFLLCKYRP 158

RESULT 11
US-09-525-041-2
; Sequence 2, Application US/09525041
; Publication No. US20030158098A1
; GENERAL INFORMATION:
; APPLICANT: Soppet et al.
; TITLE OF INVENTION: Colon Specific Gene and Protein
; FILE REFERENCE: P178D2
; CURRENT APPLICATION NUMBER: US/09/525,041
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 09/162,508
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: US 08/468,413
; PRIOR FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-525-041-2

Query Match 100.0%; Score 878; DB 10; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MASSRMRLLLLSCLAKTGVLDIIMRPSCAPGWFYHKSNCYGYFRKLKRWSDALEECQS 60
Db 1 MASSRMRLLLLSCLAKTGVLDIIMRPSCAPGWFYHKSNCYGYFRKLKRWSDALEECQS 60
Qy 61 YNGNAHLASIIISLKEASTIAEYISGYORSOPITWIGLHDPQKRQOWIDGAMLYLRWSWG 120
Db 61 YNGNAHLASIIISLKEASTIAEYISGYORSOPITWIGLHDPQKRQOWIDGAMLYLRWSWG 120
Qy 121 KSMGNKHCACEMSSNNNFLLTWSSNECNKRQHFLLCKYRP 158
Db 121 KSMGNKHCACEMSSNNNFLLTWSSNECNKRQHFLLCKYRP 158

RESULT 12
US-09-969-034-4471
; Sequence 4471, Application US/09969034
; Publication No. US20040110668A1
; GENERAL INFORMATION:

```

; APPLICANT: Burgess, Christopher C.
; APPLICANT: Aslie, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dwiwed, Poornima
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/237,271
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4471
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-969-034-4471

Query Match      100.0%; Score 878; DB 11; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MASRSMRLLLLSCLAKTGVLDIIMRPSCAPGMFYHKSNCYGFPRKLRNMSDALEECOS 60
DB      1 MASRSMRLLLLSCLAKTGVLDIIMRPSCAPGMFYHKSNCYGFPRKLRNMSDALEECOS 60

QY      61 YNGGAHLASITSLKASTIAEYISGYORSOPITWIGLHDPOKROQOWIDGAMYLYRWSWG 120
DB      61 YNGGAHLASITSLKASTIAEYISGYORSOPITWIGLHDPOKROQOWIDGAMYLYRWSWG 120

QY      121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLLCKYRP 158
DB      121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLLCKYRP 158

RESULT 13
US-10-025-380-1070
; Sequence 1070, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongrong
; APPLICANT: Jiang, Yudi
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aljun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yaeli A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1070
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1070
```

```

Query Match      100.0%; Score 878; DB 13; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MASRSMRLLLLSCLAKTGVLDIIMRPSCAPGMFYHKSNCYGFPRKLRNMSDALEECOS 60
DB      1 MASRSMRLLLLSCLAKTGVLDIIMRPSCAPGMFYHKSNCYGFPRKLRNMSDALEECOS 60

QY      61 YNGGAHLASITSLKASTIAEYISGYORSOPITWIGLHDPOKROQOWIDGAMYLYRWSWG 120
DB      61 YNGGAHLASITSLKASTIAEYISGYORSOPITWIGLHDPOKROQOWIDGAMYLYRWSWG 120

QY      121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLLCKYRP 158
DB      121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLLCKYRP 158

RESULT 14
US-10-025-380-1077
; Sequence 1077, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongrong
; APPLICANT: Jiang, Yudi
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aljun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yaeli A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1077
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1077

Query Match      100.0%; Score 878; DB 13; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MASRSMRLLLLSCLAKTGVLDIIMRPSCAPGMFYHKSNCYGFPRKLRNMSDALEECOS 60
DB      1 MASRSMRLLLLSCLAKTGVLDIIMRPSCAPGMFYHKSNCYGFPRKLRNMSDALEECOS 60

QY      61 YNGGAHLASITSLKASTIAEYISGYORSOPITWIGLHDPOKROQOWIDGAMYLYRWSWG 120
DB      61 YNGGAHLASITSLKASTIAEYISGYORSOPITWIGLHDPOKROQOWIDGAMYLYRWSWG 120

QY      121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLLCKYRP 158
DB      121 KSMGNKHCACEMSSNNNFLTWSNNECNKRQHFLLCKYRP 158

RESULT 15
US-10-025-380-1078
; Sequence 1078, Application US/10025380
; Publication No. US20020182191A1
```

```
/ GENERAL INFORMATION:
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Lodes, Michael J.
/ APPLICANT: Secrist, Heather
/ APPLICANT: Benson, Darin R.
/ APPLICANT: Meagher, Madeleine Joy
/ APPLICANT: Stolk, John A.
/ APPLICANT: Wang, Tonglong
/ APPLICANT: Jiang, Yugu
/ APPLICANT: Smith, Carole L.
/ APPLICANT: King, Gordon E.
/ APPLICANT: Wang, Aijun
/ APPLICANT: Clapper, Jonathan D.
/ APPLICANT: Skeiky, Yasir A. W.
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Vedvick Thomas S.
/ APPLICANT: Carter, Darlick
/ TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
/ TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
/ FILE REFERENCE: 210121.471C14
/ CURRENT APPLICATION NUMBER: US/10/025,380
/ NUMBER OF SEQ ID NOS: 1129
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1078
/ LENGTH: 158
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-025-380-1078

Query Match      100.0%; Score 878; DB 13; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.8e-82;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MASRSMRLLLSLSCAKTGVGDITMRPSCAPGWFYHKSNCYGRKLRNMSDALEECQS 60
DB      1 MASRSMRLLLSLSCAKTGVGDITMRPSCAPGWFYHKSNCYGRKLRNMSDALEECQS 60

QY      61 YGNGNHLASILSLKEASTIAEYISGYGRSOPITWIGLHDPKROQOWIDGAMTYLYRSMSG 120
DB      61 YGNGNHLASILSLKEASTIAEYISGYGRSOPITWIGLHDPKROQOWIDGAMTYLYRSMSG 120

QY      121 KSMGKNKCAEMSSNNNFLTWSNNECNTRQHFLECKYRP 158
DB      121 KSMGKNKCAEMSSNNNFLTWSNNECNTRQHFLECKYRP 158
```

Search completed: February 11, 2005, 23:33:33
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2005, 23:29:46 ; Search time 43 Seconds

(without alignments)
274.292 Million cell updates/sec

Title: US-09-525-041-2

Sequence: 1 MASRMRLLLLSCLAKTV.....LTWSNECKRQHPICKRP 158

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	878	100.0	158	2	US-08-729-103-1
2	878	100.0	158	2	US-08-468-413-2
3	878	100.0	158	3	US-09-162-508-2
4	878	100.0	158	5	PCT-US95-07169-2
5	695	79.2	122	4	US-09-621-976-4812
6	254.5	29.0	165	2	US-08-401-530A-7
7	254.5	29.0	165	2	US-08-729-103-3
8	254.5	29.0	165	2	US-08-709-662-7
9	240	27.3	166	2	US-08-729-103-4
10	240	27.3	174	4	US-09-949-016-10686
11	231	26.3	174	2	US-08-401-530A-5
12	231	26.3	174	2	US-08-709-662-5
13	228.5	26.0	174	2	US-08-401-530A-6
14	228.5	26.0	174	2	US-08-709-662-6
15	224	25.5	144	4	US-09-949-016-10685
16	224	25.5	166	4	US-09-949-016-6286
17	223	25.4	117	6	5514582-15
18	223	25.4	117	6	5514582-15
19	221	25.2	175	2	US-08-464-637-2
20	221	25.2	175	2	US-08-401-530A-4
21	221	25.2	175	2	US-08-709-662-4
22	221	25.2	175	2	US-08-822-261-3
23	221	25.2	175	4	US-09-226-852-3
24	221	25.2	173	4	US-09-949-016-10537
25	218	24.8	174	1	US-07-778-156-7
26	218	24.8	174	2	US-08-822-261-4
27	218	24.8	174	2	US-08-422-166-7

28	218	24.8	174	4	US-09-226-852-4	Sequence 4, Appl1
29	208	23.7	175	2	US-08-401-530A-3	Sequence 3, Appl1
30	208	23.7	175	2	US-08-709-662-3	Sequence 3, Appl1
31	204.5	23.3	130	1	US-07-893-929A-7	Sequence 7, Appl1
32	204.5	23.3	130	5	PCT-US92-10344-7	Sequence 7, Appl1
33	204	23.2	175	2	US-08-822-261-1	Sequence 1, Appl1
34	204	23.2	175	4	US-09-226-852-1	Sequence 1, Appl1
35	198.5	22.6	132	1	US-07-893-929A-5	Sequence 5, Appl1
36	198.5	22.6	132	5	PCT-US92-10344-5	Sequence 5, Appl1
37	198.5	22.6	134	1	US-07-893-929A-2	Sequence 2, Appl1
38	198.5	22.6	134	5	PCT-US92-10344-2	Sequence 2, Appl1
39	196	22.3	125	1	US-07-893-929A-3	Sequence 3, Appl1
40	196	22.3	125	5	PCT-US92-10344-3	Sequence 3, Appl1
41	194.5	22.2	131	1	US-07-893-929A-1	Sequence 1, Appl1
42	194.5	22.2	131	5	PCT-US92-10344-1	Sequence 1, Appl1
43	193	22.0	912	5	PCT-US95-03747-2	Sequence 2, Appl1
44	190	21.6	174	2	US-08-401-530A-2	Sequence 2, Appl1
45	190	21.6	174	2	US-08-709-662-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-729-103-1

Sequence 1 Application US/08729103

Patent No. 583,848

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

TITLE OF INVENTION: NOVEL HUMAN REG PROTEIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSER: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: PasteSeq Version 1.5

CURRENT APPLICATION DATA:

FILING DATE: Filed Herewith

APPLICATION NUMBER: US/08/729,103

Prior Application Number:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36, 749

REFERENCE/DOCKET NUMBER: PF-0138 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 158 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULAR TYPE: peptide

IMMEDIATE SOURCE:

LIBRARY: COLMPT02

CLONE: 1310334

US-08-729-103-1

Query Match

Best Local Similarity

Matches 158; Conservative

100.0%; Score 878; DB 2; Length 158;

Pred. No. 5.9e-88;

Mismatches 0; Indels 0; Gaps 0;


```

/ CITY: ROSELAND
/ STATE: NEW JERSEY
/ COUNTRY: USA
/ ZIP: 07068
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5 INCH DISKETTE
/ COMPUTER: IBM PS/2
/ OPERATING SYSTEM: MS-DOS
/ SOFTWARE: WORD PERFECT 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/07169
/ FILING DATE: 06 JUN 95
/ CLASSIFICATION:
/ PRIORITY APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: FERRARO, GREGORY D.
/ REGISTRATION NUMBER: 36,134
/ REFERENCE/DOCKET NUMBER: 325800-389
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201-994-1700
/ TELEFAX: 201-994-1744
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 158 AMINO ACIDS
/ TYPE: AMINO ACID
/ STRANDEDNESS:
/ TOPOLOGY: LINEAR
/ MOLECULE TYPE: PROTEIN
/ PCT-US95-07169-2

Query Match      100.0%; Score 878; DB 5; Length 158;
Best Local Similarity 100.0%; Pred. No. 5.9e-88;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASRSRRLLLSLCLAKTGVLDIIMRSCAGMFWYKSNCGYFPRKLNMSDALEECOS 60
DB 1 MASRSRRLLLSLCLAKTGVLDIIMRSCAGMFWYKSNCGYFPRKLNMSDALEECOS 60

QY 61 YNGGAHLASILSLKEASTIAEYISGYQSRQPIWIGLHDPQKQOMWIDGAMVLYRSMG 120
DB 61 YNGGAHLASILSLKEASTIAEYISGYQSRQPIWIGLHDPQKQOMWIDGAMVLYRSMG 120

QY 121 KSMGKNCACWSSNNNFLTWSNNECNKQHFCKYRP 158
DB 121 KSMGKNCACWSSNNNFLTWSNNECNKQHFCKYRP 158

RESULT 5
US-09-621-976-4812
/ Sequence 4812, Application US/09621976
/ Patent No. 6639063
/ GENERAL INFORMATION:
/ APPLICANT: Dumas Milne Edwards, J.B.
/ APPLICANT: Jobert, S.
/ APPLICANT: Giordano, J.Y.
/ TITLE OF INVENTION: ESTs and Encoded Human Proteins.
/ FILE REFERENCE: GENSET. 054PR2
/ CURRENT APPLICATION NUMBER: US/09/621, 976
/ CURRENT FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 19335
/ SOFTWARE: Patent.pm
/ SEQ ID NO 4812
/ LENGTH: 122
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-621-976-4812

Query Match      79.2%; Score 695; DB 4; Length 122;
Best Local Similarity 100.0%; Pred. No. 3.9e-68;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 26 MRSCAPGMFWYKSNCGYFPRKLNMSDALEECOSYNGGAHLASILSLKEASTIAEYISG 85
DB 1 MRSCAPGMFWYKSNCGYFPRKLNMSDALEECOSYNGGAHLASILSLKEASTIAEYISG 60

QY 86 YORSQPIWIGLHDPQKQOMWIDGAMVLYRSMGSKSGKNGKCAWSSNNNFLTWSNNE 145
DB 61 YORSQPIWIGLHDPQKQOMWIDGAMVLYRSMGSKSGKNGKCAWSSNNNFLTWSNNE 120

QY 146 CN 147
DB 121 CN 122

RESULT 6
US-08-401-530A-7
/ Sequence 7, Application US/08401530A
/ Patent No. 5834590
/ GENERAL INFORMATION:
/ APPLICANT: Vink, Aaron I.
/ APPLICANT: Pittenger, Gary L.
/ APPLICANT: Rafaeloff, Ronit
/ APPLICANT: Rosenberg, Lawrence
/ APPLICANT: Duguid, William P.
/ TITLE OF INVENTION: INGP PROTEIN INVOLVED IN PANCREATIC
/ NUMBER OF SEQUENCES: 7
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Banner & Allegretti
/ STREET: 1001 G Street, N.W.
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: US
/ ZIP: 20001-4597
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/401, 530A
/ FILING DATE: 22-FEB-1995
/ CLASSIFICATION: 800
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Kagan, Sarah A.
/ REGISTRATION NUMBER: 32,141
/ REFERENCE/DOCKET NUMBER: 00570.48743
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-508-9100
/ TELEFAX: 202-508-9299
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 165 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: Protein
/ ORIGINAL SOURCE:
/ ORGANISM: Rattus rattus
US-08-401-530A-7

Query Match      29.0%; Score 254.5; DB 2; Length 165;
Best Local Similarity 30.6%; Pred. No. 6.6e-20;
Matches 52; Conservative 38; Mismatches 51; Indels 29; Gaps 7;

QY 7 RLLLLSLCLAKTGVLDIIMRPS-----CAPGMFWYKSNCGYFPRKLNW 51
DB 5 KYFILLSC-----WVLSPSQCGAEEDLPARITCPREGSNAYSSCYCYFMDHLSW 56

QY 52 SDARECQSYNGGAHLASILSLKEASTIAEYI--SGYQSRQPIWIGLHDPQKQOMWID 109
DB 57 AADLPFCQNMWSG-VLAVSVLSQABGNPLASLIKESGTTAA-N-VWIGLHDPKNNRWMS 114

QY 110 GAMVLYRSM-SGKSGKNGK-HCAWSSNNNFLTWSNNECNKQHFCKYRP 157
DB 110 GAMVLYRSM-SGKSGKNGK-HCAWSSNNNFLTWSNNECNKQHFCKYRP 157
```


MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/729,103
FILING DATE: Filed Herewith
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0138 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 166 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 474306
US-08-729-103-4

Query Match 27.3%; Score 240; DB 2; Length 166;

Best Local Similarity 34.3%; Pred. No. 2.6e-18; Matches 46; Conservative 27; Mismatches 57; Indels 4; Gaps 4;

QY 27 RPSCAPGFYHNSKCYGFRKLRNMSDALEECOSYNGAHASISLKEASTIAEYISGY 86
DB 33 RISCEGTNAVRSYCYFNEDEPETWADLYCQNNNSG-NLVSVLQAEGAFVSLIKES 91
QY 87 QRSQ-ITWGLHDPKQKQOMWIDGAMVLYRSW-SGKSMGNK-HCAEMSSNNPLTWSS 143
DB 92 STDSDSNVWIGLHDPKKNRMRHMSGSLVSYKSWDTGSPSANAGYCASLTCGFKKWD 151
QY 144 NECNRQHFCLKYR 157
DB 152 ESECKKFSFVCKFK 165

RESULT 10

US-09-949-016-10686
Sequence 10686, Application US/09949016

Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CL001307

CURRENT APPLICATION NUMBER: US/09/949,016

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10686

LENGTH: 174

TYPE: PRT

ORGANISM: Human

US-09-949-016-10686

Query Match 27.3%; Score 240; DB 4; Length 174;
Best Local Similarity 34.3%; Pred. No. 2.7e-18;

Matches 46; Conservative 27; Mismatches 57; Indels 4; Gaps 4;
QY 27 RPSCAPGFYHNSKCYGFRKLRNMSDALEECOSYNGAHASISLKEASTIAEYISGY 86
DB 41 RISCEGTNAVRSYCYFNEDEPETWADLYCQNNNSG-NLVSVLQAEGAFVSLIKES 99
QY 87 QRSQ-ITWGLHDPKQKQOMWIDGAMVLYRSW-SGKSMGNK-HCAEMSSNNPLTWSS 143
DB 100 STDSDSNVWIGLHDPKKNRMRHMSGSLVSYKSWDTGSPSANAGYCASLTCGFKKWD 159
QY 144 NECNRQHFCLKYR 157
DB 160 ESECKKFSFVCKFK 173

RESULT 11

US-08-401-530A-5

Sequence 5, Application US/08401530A

Patent No. 5814590

GENERAL INFORMATION:

APPLICANT: Vinik, Aaron I.

APPLICANT: Piltenger, Gary L.

APPLICANT: Rafaeloff, Ronit

APPLICANT: Rosenberg, Lawrence

APPLICANT: Duguid, William P.

TITLE OF INVENTION: IN GAP PROTEIN INVOLVED IN PANCREATIC

NUMBER OF SEQUENCES: 7

TITLE OF INVENTION: ISLET NEOGENESIS

CORRESPONDENCE ADDRESS:

STREET: 1001 G Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: US

ZIP: 20001-4597

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/401,530A

FILING DATE: 22-Feb-1995

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 00570.48743

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 174 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Rattus rattus

US-08-401-530A-5

Query Match 26.3%; Score 231; DB 2; Length 174;

Best Local Similarity 32.4%; Pred. No. 2.6e-17;

Matches 55; Conservative 28; Mismatches 73; Indels 14; Gaps 5;

QY 1 MASRSMRLILLISCLAKTGVGD-----IMRPSCAPGFYHNSKCYGFRKLRNMSDA 54
DB 5 VALTMSWMLLSLMSLISQVQGBDAKEDVPTSRISCPKSGRAYGYCYALPSVSKSWFDA 64
QY 55 BLECOSYNGAHASISLKEASTIAEYI-SGYORSQPIWIGLHDPKQKQ-----WQMI 108
DB 65 DLACQKRSQ-NLVSVLQSGEASFSVSLIKSGNCGQVWVIGLHDPKQKQ-----WQMI 123

STREET: 1001 G Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: US
ZIP: 20001-4597
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/709,662
FILING DATE: 09-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 00570.59178
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Rattus rattus
US-08-709-662-6

Query Match 26.0%; Score 228.5; DB 2; Length 174;
Best Local Similarity 34.5%; Pred. No. 4, 9e-17;
Matches 48; Conservative 25; Mismatches 57; Indels 9; Gaps 4;

QY 27 RPSCAPGFYHKSNCYGRKLRNMSDAELBCOSYNGAHLASILSKASTIAEYISG- 85
DB 36 RISCPEGTNAYRSYCYTYNEDRETVDADLVQNNNSG-NLVSVLTQAEAFVASLIKES 69
QY 86 YGRSQPIWIGLHDPQKROQ-----WQIDGAMLYRSWSG--KSMGKNGKCAEMSSNNF 138
DB 95 VNNNDIWIWIGLHDPQKROQNGGKMSNSDVNIYLMNDPSSYVNRGCGSLTATSEF 154
QY 139 LTWSNECKRQHFLCKYR 157
DB 155 LKWGDHCHDVELPFVCKFK 173

RESULT 15
US-09-949-016-10685
Sequence 10685, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10685
LENGTH: 144
TYPE: PRT
ORGANISM: Human
US-09-949-016-10685

Query Match 25.5%; Score 224; DB 4; Length 144;
Best Local Similarity 33.3%; Pred. No. 1, 2e-16;
Matches 45; Conservative 26; Mismatches 58; Indels 6; Gaps 4;

QY 27 RPSCAPGFYHKSNCYGRKLRNMSDAELBCOSYNGAHLASILSKASTIAEYISG- 84
DB 11 RISCPEGTNAYRSYCYTYNEDRETVDADLVQNNNSG-NLVSVLTQAEAFVASLIKES 69
QY 85 YGRSQPIWIGLHDPQKROQWQIDGAMLYRSW--KSMGKNGKCAEMSSNNF 142
DB 70 GTDDFN-VWIGLHDPKRRMHTSSGSLVSKWIGAPSSVNPQCVSLTSSGTFORWK 128
QY 143 SNECKRQHFLCKYR 157
DB 129 DVPCEDKPSFVCKFK 143

Search completed: February 11, 2005, 23:43:03
Job time : 44 secs

This Page Blank (uspto)